

**ÉCOLE DOCTORALE Informatique, Télécommunications et Électronique de  
Paris (ED130)**

**Centre d'Études et de Recherche en Informatique et Communications (CEDRIC)  
Conception et Programmation Raisonnées (CPR)  
Vérification et Évaluation de Systèmes Parallèles et Asynchrones (VESPA)**

**THÈSE** présentée par :

**Jérôme DANTAN**

soutenue le : 1<sup>er</sup> juin 2016

pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Discipline/ Spécialité : Informatique

**Une approche systémique unifiée pour  
l'optimisation durable des systèmes socio-  
environnementaux**

**Ingénierie des systèmes de décision en univers  
incertain**

**THÈSE dirigée par :**

**M. POLLET Yann**

Professeur Titulaire de la Chaire d'Intégration des Systèmes,  
CNAM Paris

**THÈSE co-encadrée par :**

**Mme TAÏBI Salima**

Responsable du pôle modélisation, HDR, UniLaSalle

**RAPPORTEURS :**

**M. BONJOUR Eric**

Professeur des Universités, Université de Lorraine

**M. CARDON Alain**

Professeur des Universités, Université du Havre

---

**JURY :**

**M. BONJOUR Eric**

Professeur des Universités, Université de Lorraine

**M. CARDON Alain**

Professeur des Universités, Université du Havre

**M. ITMI Mhamed**

Maître de Conférences, HDR, INSA Rouen

**M. KOUTANI Saïd**

Directeur des Formations, UniLaSalle

**Mme LEVY Nicole**

Professeur des Universités, CNAM Paris

**M. POLLET Yann**

Professeur Titulaire de la Chaire d'Intégration des Systèmes,  
CNAM Paris

**Mme TAÏBI Salima**

Responsable du pôle modélisation, HDR, UniLaSalle



## Remerciements

Je tiens tout d'abord à remercier Yann Pollet, mon directeur de thèse. Son encadrement, ses conseils, son aide, sa patience et son soutien constant ont été pour moi au cours de ces dernières années source de motivation, d'encouragement, de réconfort et de persévérance.

Je tiens également à remercier Salima Taïbi, qui a également participé à l'encadrement de ma thèse, pour ses conseils avisés et son soutien.

Je remercie les membres du jury pour l'intérêt qu'ils manifestent à l'égard de ce travail. Je tiens à exprimer ma profonde gratitude envers Éric Bonjour et Alain Cardon pour l'honneur qu'ils me font d'accepter d'être les rapporteurs de cette thèse. Je tiens à remercier vivement Mhamed Itmi, Nicole Lévy et Saïd Koutani d'avoir accepté de faire partie du jury de cette thèse.

Mes remerciements s'adressent également à mes collègues d'UniLaSalle qui m'ont fait profiter de leur soutien et de leurs conseils et plus particulièrement Patrice Lepelletier.

Je tiens également à exprimer ma reconnaissance envers Karine Laval et Saïd Koutani qui m'ont permis de réaliser ce travail dans les meilleures conditions.

Enfin je remercie ma famille qui m'a toujours soutenu : ma femme Nathalie, notre fils Matthieu ainsi que mes parents Claire et Alain Dantan.

# Résumé

De nos jours, la durabilité des activités humaines devient une préoccupation majeure dans le monde entier. Il s'agit d'évaluer ces activités non seulement en matière d'efficacité et de productivité, mais aussi en ce qui concerne leurs durabilités économique, sociale, environnementale, etc. Pour ce faire, les experts de ces différents domaines doivent travailler en collaboration.

Dans ce contexte, les sociétés humaines sont confrontées à plusieurs défis majeurs qui sont les suivants : (1) traiter de grandes quantités d'informations, dont le volume augmente de manière exponentielle (« big data »), (2) évoluer dans un monde réel dynamique et imparfait : les sources de données fournissent en effet des représentations à la fois partielles, imparfaites et immédiates du monde et de ses activités en temps réel, (3) prévoir et évaluer les états futurs de ses activités en raison de la complexité des règles régissant leur comportement et leur évolution : rétroactions, couplages, interactions et sources d'informations multiples.

Les recherches que nous avons menées dans cette thèse contribuent plus particulièrement au domaine de l'ingénierie des systèmes de décision en univers incertain, dans un contexte de grandes masses de données traitées dynamiquement grâce à des modèles complexes résultant de l'assemblage de modèles élémentaires. Nous prenons comme objet d'étude général le domaine des systèmes socio-environnementaux, et plus particulièrement le domaine pluridisciplinaire de l'agriculture.

Nous proposons une approche systémique pour l'optimisation durable des systèmes socio-environnementaux. Notre approche unifie des apports des différents domaines concernés, tirant parti des avantages de chacun d'entre eux et les faisant opérer en synergie : (1) la méta-modélisation des systèmes socio-environnementaux via un méta-modèle SysML, (2) la représentation générique de l'imperfection des informations qui circulent dans ces systèmes, associée à un modèle de décision en contexte incertain et enfin (3) la simulation et l'évaluation de ces systèmes en environnement dynamique en vue de prises de décisions par des experts, que nous avons illustrée par un modèle d'architecture orientée service ainsi que des études de cas appliquées au domaine de l'agriculture.

Mots-clés : décision, incertitude, optimisation durable, possibilité, probabilité, systèmes socio-environnementaux.

# Abstract

Nowadays, the sustainability of human activities is a major worldwide concern. The challenge is to evaluate such activities not only in terms of efficiency and productivity, but also in terms of their economic, social, environmental, etc. durability. For this, the experts of these areas need to work collaboratively.

In this context, human societies are facing several major challenges such as: (1) process a large amount of information whose volume increases exponentially ("big data"), (2) live in a both dynamic and imperfect real world: data sources indeed provide partial, imperfect and immediate representations of both the real world and its activities in real time, (3) predict and assess future states of its activities due to the complexity of rules governing both their behavior and changes: feedbacks, couplings, interactions and multiple sources of information.

The researches we have conducted in this thesis contribute in particular to the domain of decision systems engineering under uncertainty in a context of a large amount of data that are dynamically processed through complex patterns, resulting from the assembly of elementary models. We have chosen the field of general socio-environmental systems as subject of study, particularly the multidisciplinary field of agriculture.

We propose a systemic approach for the sustainable optimization of socio-environmental systems. Our approach unifies the contributions of various areas concerned, leveraging the benefits of each of them and making them operate in synergy: (1) the meta-modeling of socio-environmental systems via a SysML meta-model, (2) the generic representation of data imperfection flowing in such systems, associated to a decision model in uncertain environment and finally (3) the simulation and the assessment of such systems in dynamic environment for the purpose of decision making by experts which we have illustrated by both a service-oriented architecture model and case studies applied to the agriculture domain.

Keywords: decision, uncertainty, sustainable optimization, possibility, probability, socio-environmental system.

# Table des matières

Remerciements .....	3
Résumé .....	4
Abstract .....	5
Table des matières .....	6
Liste des tableaux .....	12
Liste des figures .....	14
Liste des annexes.....	16
Première partie : introduction générale .....	17
Contexte .....	18
Problématique.....	20
Approche .....	23
Contributions .....	25
Deuxième partie : état de l’art .....	28
1    Chapitre 1 : l’approche analytique : modélisation de fonctions de productions .....	29
1.1    Bases théoriques .....	29
1.2    Applications aux systèmes socio-environnementaux.....	31
1.3    Limites de l’approche analytique .....	33
2    Chapitre 2 : l’approche systémique : modélisation et simulation de systèmes .....	34
2.1    Les modèles systémiques existants pour la simulation de systèmes socio-environnementaux .....	35
2.2    L’approche systémique .....	38
2.3    SysML, le langage standard de modélisation des systèmes .....	39
2.3.1    Vue générale.....	39
2.3.2    Modélisation de la structure d'un système socio-environnemental.....	44

2.3.3	Modélisation de la structure des relations entre les éléments d'un système .....	47
2.3.4	Modélisation des relations numériques et des contraintes sur les valeurs des paramètres d'un système .....	50
2.4	Simulation de systèmes .....	52
2.5	Conclusion : les limites des approches actuelles de modélisation et simulation des systèmes .....	53
3	Chapitre 3 : aide à la décision, modélisation et manipulation d'informations imparfaites	55
3.1	Aide à la décision .....	56
3.1.1	Choix et préférences .....	56
3.1.2	Les fonctions d'utilité.....	57
3.1.3	Décisions en contexte incertain.....	58
3.1.4	Analyse de la décision dans le risque.....	59
3.1.5	Analyse de la décision en contexte incertain .....	61
3.1.6	Agrégation de décisions multicritères .....	63
3.1.7	Décisions à partir d'informations imparfaitement définies.....	65
3.1.8	Applications dans le domaine de l'agriculture .....	70
3.1.9	Conclusion et limites des approches d'aide à la décision .....	74
3.2	Modélisation et manipulation d'informations imparfaites .....	75
3.2.1	Introduction .....	75
3.2.2	L'analyse de l'imperfection par les mesures de confiance .....	77
3.2.3	Modélisation de phénomènes aléatoires.....	78
3.2.4	Modélisation d'informations incertaines, incomplètes, imprécises, linguistiques	79
3.2.5	Modélisation de connaissances à la fois aléatoires et floues : les « fuzzy random variables » .....	83
3.2.6	Combinaison d'informations incertaines, incomplètes, imprécises, linguistiques et aléatoires : la théorie de Dempster-Shafer .....	84

3.2.7	Combinaison de variables aléatoires et de distributions de possibilité par encadrement de fonctions de répartition .....	87
3.3	Conclusion : les limites des approches classiques de modélisation et de combinaison d'informations imparfaites .....	88
	Troisième partie : contributions .....	90
4	Chapitre 4 : un méta-modèle systémique pour l'optimisation durable des systèmes socio-environnementaux .....	91
4.1.1	Un méta-modèle systémique dédié aux systèmes socio-environnementaux.....	92
4.1.2	Modélisation du domaine : un modèle générique d'exploitation agricole .....	94
4.1.3	Un modèle structurel définissant les différents types d'imperfection des données rencontrés dans les systèmes socio-environnementaux .....	96
4.1.4	Le modèle fonctionnel.....	100
4.1.5	Intégration des nouveaux concepts au langage de modélisation standard des systèmes SysML.....	102
4.1.6	Conclusion : un méta-modèle systémique nécessitant un formalisme unifié pour représenter et opérer sur l'information imparfaite .....	112
5	Chapitre 5 : un formalisme unifié pour représenter et opérer sur l'information imparfaite	113
5.1	Introduction .....	113
5.2	De l'utilité de distinguer probabilités et possibilités dans un contexte de prise de décision.....	114
5.3	Formalisation et calculs génériques sur des connaissances probabilistes et possibilistes discrètes .....	116
5.3.1	Notations et définitions .....	116
5.3.2	Quantités possibilistes discrètes .....	117
5.3.3	Quantités probabilistes discrètes .....	123
5.3.4	Quantités mixtes discrètes .....	129



5.4	Formalisation et calculs génériques sur des connaissances probabilistes et possibilistes continues .....	133
5.4.1	Quantités possibilistes continues.....	135
5.4.2	Quantités probabilistes continues.....	140
5.4.3	Quantités mixtes continues.....	144
5.4.4	Cas particuliers .....	145
5.5	Structure informatique : modélisation de l'imperfection dans des bases de données 146	
5.5.1	Modélisation « nette » .....	146
5.5.2	Modélisation de l'imperfection .....	147
5.5.3	Opérateur de sélection .....	151
5.5.4	Opérateurs d'agrégation .....	154
5.6	Conclusion : un formalisme unifié de gestion de connaissances imparfaites .....	157
6	Chapitre 6 : un modèle de décision en contexte incertain.....	159
6.1	Des interprétations particulières de combinaisons de valeurs entachées d'incertitudes .....	160
6.1.1	Une première interprétation particulière des incertitudes sur les valeurs : l'approche Dempster-Shafer, combinaison conjonctive de valeurs probabilistes et possibilistes .....	160
6.1.2	Une seconde interprétation particulière : transformation normalisation à un en distribution de possibilité .....	164
6.1.3	Une première interprétation particulière des variations de valeurs : approche par transformation de distributions de possibilité en famille de probabilités.....	166
6.1.4	Une seconde interprétation particulière des variations de valeurs : approche par transformation de distributions de probabilité en distributions de possibilité .....	168
6.1.5	Cas particuliers : combinaisons du cas continu et du cas discret .....	168
6.2	Des opérateurs génériques d'agrégation .....	169

6.2.1	Distributions de probabilité : cas discret .....	170
6.2.2	Distributions de probabilité : cas continu.....	171
6.2.3	Distributions de possibilité : cas discret .....	172
6.2.4	Distributions de possibilité : cas continu.....	174
6.3	Conclusion : des interprétations de combinaisons des distributions incertaines, des opérateurs génériques d'agrégation pour la prise de décision en contexte incertain .....	175
Quatrième partie : validation.....		177
7	Chapitre 7 : évaluation : des scénarios d'aide à la décision sur la période de semis dans une exploitation agricole .....	178
7.1	Étude de cas 1 : de l'utilité de distinguer probabilités et possibilités .....	178
7.1.1	Cas 1 : distributions de possibilité discrètes et valeurs à la fois précises et certaines.....	180
7.1.2	Cas 2 : distributions de possibilité discrètes et ensembles des valeurs aléatoires	181
7.1.3	Cas 3 : distributions de possibilité continues et ensembles des valeurs aléatoires	184
7.2	Étude de cas 2 : cas de probabilités et possibilités imbriquées .....	187
7.2.1	Interprétation 1 : mesures de Dempster-Shafer.....	188
7.2.2	Interprétation 2 : normalisation à un .....	192
7.3	Conclusion : deux études de cas pour l'aide à la décision dans des exploitations agricoles mettant en œuvre les concepts précédemment définis.....	195
8	Chapitre 8 : opérationnalisation : un modèle architectural pour la mise en œuvre d'applications dédiées aux systèmes socio-environnementaux .....	196
8.1	Introduction : architectures orientées services et services Web .....	196
8.2	Processus de conception d'une application innovante dédiée aux systèmes socio-environnementaux .....	198
8.3	Une architecture orientée service dédiée aux systèmes socio-environnementaux.	200

8.3.1	Services externes .....	200
8.3.2	Interface utilisateur.....	200
8.3.3	Illustration de l'application proposée.....	201
8.3.4	Index des services exposés .....	203
8.4	Fonctions proposées .....	205
8.4.1	Une interface pour la gestion de l'imperfection des données .....	205
8.4.2	Une interface pour l'introspection des fonctions et classes .....	206
8.5	Conclusion : une nouvelle approche de conception d'applications innovantes dédiées aux systèmes socio-environnementaux .....	206
Cinquième partie : conclusion.....		207
Conclusion générale et perspectives .....		208
Synthèse des travaux réalisés .....		208
Apports au domaine de l'ingénierie des systèmes socio-environnementaux .....		209
Perspectives .....		211
Publications et conférences .....		212
Publication réalisée dans le cadre de la thèse.....		212
Communications avec actes dans des conférences d'audience internationale réalisées dans le cadre de la thèse .....		212
Autres publications et conférences.....		213
Publication.....		213
Communications.....		213
Autres productions .....		215
9	Bibliographie .....	217
Résumé.....		242
Résumé en anglais .....		242

## Liste des tableaux

Tableau 1 : diagrammes dynamiques et comportementaux de SysML.....	41
Tableau 2 : diagrammes structurels de SysML .....	42
Tableau 3 : matrice de décision.....	59
Tableau 4 : tableau des rangs par critère de décision.....	63
Tableau 5 : t-normes et t-conormes les plus utilisées (source : [Bouchon-Meunier, 1993]) ...	66
Tableau 6 : valeurs de vérité des différents types d'implication floue (source : [Bouchon-Meunier, 1993]).....	67
Tableau 7 : types d'incertitudes de l'espérance et de la variance pour les différentes FRVs [Shapiro, 2012].....	84
Tableau 8 : l'ADT « fuzzyDiscrete » .....	98
Tableau 9 : l'ADT trapezoid .....	98
Tableau 10 : l'ADT probableDiscrete .....	99
Tableau 11 : l'ADT normal.....	100
Tableau 12 : modèle du domaine et fonctions spécifiques à l'agriculture .....	103
Tableau 13 : possibilités et nécessités croisées de la base $B\pi$ avec l'objet $O\pi$ .....	153
Tableau 14 : probabilités conjointes de la base $Bp$ avec l'objet $Op$ .....	154
Tableau 15 : valeurs possibles des opérateurs d'agrégation classiques pour le cas possibiliste .....	155
Tableau 16 : valeurs possibles des opérateurs d'agrégation classiques pour le cas probabiliste .....	156
Tableau 17 : opérateurs sur les quantités probabilistes discrètes (1) .....	170
Tableau 18 : opérateurs sur les quantités probabilistes discrètes (2) .....	170
Tableau 19 : opérateurs sur les quantités probabilistes continues (1) .....	171
Tableau 20 : opérateurs sur les quantités probabilistes continues (2) .....	171
Tableau 21 : opérateurs sur les quantités possibilistes discrètes (1) .....	173

Tableau 22 : opérateurs sur les quantités possibilistes discrètes (2) .....	173
Tableau 23 : opérateurs sur les quantités possibilistes continues (1) .....	174
Tableau 24 : opérateurs sur les quantités possibilistes discrètes (2) .....	175
Tableau 25 : récapitulatif des deux scénarios possibles .....	179
Tableau 26 : cas 1 : utilités des différents scénarios .....	180
Tableau 27 : cas 2 : utilités des différents scénarios .....	182
Tableau 28 : cas 3 : lien profondeur de semis / nombre flou .....	184
Tableau 29 : synthèse des décisions calculées en fonction des critères et des cas.....	187
Tableau 30 : synthèse des différents résultats obtenus en fonction des interprétations réalisées pour chaque scénario .....	194
Tableau 31 : synthèse des décisions calculées en fonction des critères et des interprétations	194
Tableau 32 : ressources du modèle structurel .....	203
Tableau 33 : ressources du modèle fonctionnel .....	203
Tableau 34 : ressources du modèle opérationnel .....	204

# Liste des figures

Figure 1 : vue d'ensemble des étapes du processus de KDD (source : [Fayyad, U. et al, 1996]) .....	30
Figure 2 : principe de fonctionnement de STICS.....	37
Figure 3 : les différents aspects de la modélisation d'un système socio-environnemental.....	39
Figure 4 : les différents diagrammes SysML (source : [OMG, 2015]) .....	41
Figure 5 : les quatre piliers de SysML (source : [OMG, 2015]) .....	43
Figure 6 : exemple de diagramme de définition de bloc modélisant un agrosystème .....	47
Figure 7 : exemple de diagramme de bloc interne modélisant un agrosystème.....	50
Figure 8 : diagramme paramétrique représentant les différentes équations du modèle de Hénin-Dupuis (légende des unités : kg = kilogramme ; a = année ; ha : hectare).....	52
Figure 9 : arborescence hiérarchique de la contribution au développement durable dans le modèle MASC 2.0 (source : [Craheix et al., 2012]) .....	72
Figure 10 : exemple de matrice croisant les objectifs avec les indicateurs (source : [Vilain et al., 2003]) .....	73
Figure 11 : échelle de durabilité agro-écologique subdivisée en indicateurs (source : [Vilain et al., 2003]) .....	74
Figure 12 : différents types d'imperfection des données .....	76
Figure 13 : relations entre alpha coupes et masses focales .....	86
Figure 14 : le modèle général en « diamant » .....	91
Figure 15 : le méta-modèle systémique - liens avec le modèle général .....	93
Figure 16 : le modèle générique d'exploitation agricole.....	95
Figure 17 : le modèle des entités.....	100
Figure 18 : schéma récapitulatif du modèle fonctionnel .....	102
Figure 19 : diagramme de définition de bloc de l'exploitation agricole intégrant l'imperfection des données (cycle d'un fertilisant).....	107

Figure 20 : diagramme de bloc interne de l'exploitation agricole intégrant l'imperfection des données (cycle d'un fertilisant) .....	109
Figure 21 : diagramme paramétrique de l'exploitation agricole intégrant l'imperfection des données (cycle d'un fertilisant).....	111
Figure 22 : famille de probabilités induite par une distribution de possibilité.....	167
Figure 23 : architecture de l'application proposée .....	202

## Liste des annexes

Annexes .....	228
Annexe : code source de l'application proposée .....	229



Première partie :  
introduction générale

# Contexte

De nos jours, la durabilité des activités humaines, telle que celle des entreprises, devient une préoccupation majeure dans le monde entier. En effet, il ne s'agit plus d'évaluer seulement l'efficacité, mais aussi la durabilité selon différents axes qui peuvent être de diverses natures : économiques, sociaux, environnementaux, etc. Les buts de toute entreprise humaine sont donc de pérenniser et d'optimiser durablement ses activités selon plusieurs axes d'amélioration. L'atteinte de ces buts constitue un défi majeur pour la société.

Ces notions de pérennisation multi-aspects et multidisciplinaire font que les experts de différents domaines doivent travailler de manière collaborative. Avec l'avènement des technologies de l'information et de la communication, chaque domaine peut potentiellement contribuer aux travaux d'autres domaines.

Les activités humaines agissent aussi bien sur des éléments naturels (e.g. les ressources de la planète), que sur des éléments artificiels mis en œuvre par les acteurs du domaine, et destinés à exploiter ou contrôler les premiers. Avec l'apparition de l'industrie de masse puis de la société de l'information, les règles régissant le comportement et l'évolution de ces activités deviennent de plus en plus compliquées à définir, si bien que leurs états futurs sont difficiles à prévoir et à évaluer. Des rétroactions, des couplages, des interactions mais aussi des multiples sources d'informations issues de capteurs (drones, satellites, mesures en temps réel) contribuent à cette complexité.

Dans ce contexte, un autre défi majeur est de traiter de grandes quantités d'informations, dont le volume augmente de manière exponentielle, notamment grâce au développement des moyens de génération de telles informations, par exemple les objets intelligents, les capteurs, l'émergence de l'open data, les réseaux sociaux, etc., et aussi grâce au développement des moyens de stockage. Les moyens de stockage et de traitement de grandes quantités de données sont communément réunis sous l'expression en vogue « big data ».

Enfin, les activités humaines évoluent dans un monde dynamique qui est en évolution permanente : des crises, des catastrophes, des changements climatiques perturbent ces activités qui évoluent donc en contexte incertain. De plus, les technologies telles que les capteurs et objets connectés contribuent à ce dynamisme : ils fournissent en effet des

représentations à la fois partielles, imparfaites et immédiates du monde et de ses activités en temps réel.

Nous prenons comme objet d'étude général le domaine des systèmes socio-environnementaux, et plus particulièrement le domaine pluridisciplinaire de l'agriculture.

# Problématique

Dans ce paragraphe, nous présenterons la problématique générale abordée dans la présente thèse. Les points définissant plus précisément cette problématique consistent à évaluer les activités humaines dans le but de leur optimisation durable<sup>1</sup> dans le cadre des systèmes socio-environnementaux, en environnement incertain.

Dans ce contexte, nous sommes confrontés aux problèmes suivants :

- La nécessité de traiter des grandes masses de données brutes (aspect « volume ») de l'ordre de zettaoctets<sup>2</sup> venant de sources multiples ;
- L'exploration de données de formats hétérogènes d'un point de vue aussi bien technique (formats de données) que mathématique (distributions des données, imperfections). D'une part, ces hétérogénéités sont liées aux multiples sources d'informations entraînant une grande variété des formats de données, intensifiée par le caractère multidisciplinaire des domaines en jeu. D'autre part, les nouvelles technologies en jeu (capteurs, drones) ainsi que l'aspect temps réel sans optimisation des échantillons sont sources d'imperfections sur les informations récoltées (impact sur leur qualité des données).
- Ces données ont un caractère multidimensionnel<sup>3</sup> - de par leurs niveaux d'imprécision, leurs caractères multi-échelles, leur quantité, etc. - et peuvent être issues de variables pouvant être corrélées. En effet, ces dimensions multiples sont fortement imbriquées et il existe donc de nombreux couplages (i.e. des dépendances, des rétroactions, etc.) entre les différents aspects de ces systèmes (complexité). Ces couplages peuvent

---

<sup>1</sup> Nous entendons ici par « optimisation durable » un concept d'optimisation prenant en compte un certain nombre de facteurs de projection dans l'avenir, ne se limitant pas à l'optimisation du seul état présent d'un système.

<sup>2</sup> Unité de mesure de la quantité d'information numérique valant  $10^{21}$  octets (symbole : Zo).

<sup>3</sup> Au sens mathématique du terme : données composées d'éléments d'espaces vectoriels à plusieurs dimensions

entraîner une concurrence entre critères d'évaluation et de décision dans les systèmes socio-environnementaux. En effet, il peut être difficile de concilier des critères comme « gagner de l'argent » ou bien « être productif » avec d'autres tels que « créer du bien-être » ou « préserver la nature ».

- Les données en jeu peuvent être statiques, c'est-à-dire « nettoyées » puis stockées dans des bases de données, en attendant une exploitation ultérieure. Cependant, une difficulté particulière concerne l'aspect dynamique de certaines informations. Celles-ci concernent des événements telles que mises à jour et aléas ayant un impact immédiat sur notre vue du monde réel ; elles doivent donc être traitées en temps réel sans forcément être préalablement stockées en base (« vélocité ») ; le contexte technologique dans lequel nous nous situons permet la génération de telles informations.

Ces problèmes nécessitent de développer des approches méthodologiques adaptées telles que les interventions d'experts de multiples domaines pour fournir des données et des modèles analytiques, systémiques et d'évaluation adaptés en vue de prendre des décisions, participant ainsi à la collaboration entre ces domaines.

En effet, les besoins des acteurs des différentes activités liées aux systèmes socio-environnementaux (inter et intra entreprises, experts, agriculteurs, collectivités, etc.) sont les suivants :

- Le traitement au fil de l'eau d'informations issues de différentes sources sans forcément les stocker au préalable.
- L'évaluation et l'optimisation de ces activités selon de multiples aspects, points de vue, et critères, plus particulièrement dans le cadre des systèmes socio-environnementaux qui sont pluridisciplinaires.
- La prise de décisions en termes de stratégie, de management et de gestion agricoles, à l'aide d'indicateurs d'aide à la décision en contexte dynamique, sans forcément avoir le temps nécessaire pour en vérifier la qualité et optimiser les échantillons.
- Le traitement, l'évaluation et la prise de décision précédemment abordées se font en contexte incertain de par l'imperfection de l'information et du caractère hypothétique d'événements passés ou futurs.
- Enfin, le contexte économique actuel nécessite une optimisation des ressources et des gains de temps entraînant une diminution des coûts, notamment par l'automatisation

de certains processus : d'une part dans la conception des modèles de prévision et de prise de décision et d'autre part dans le recueil des données, en évitant de construire de coûteux recueils de données spécifiques, par exemple en agrégeant celles-ci grâce à des méta-analyses scientifiques.

# Approche

Les recherches que nous avons menées dans cette thèse contribuent plus particulièrement au domaine de l'ingénierie des systèmes de décision en univers incertain, dans un contexte de grandes masses de données traitées dynamiquement par des modèles complexes résultant de l'assemblage de modèles élémentaires.

Nous proposons donc une approche unifiant des apports des différents domaines concernés, tirant parti des avantages de chacun d'entre eux et les faisant opérer en synergie : la modélisation des systèmes via un méta-modèle pour les systèmes socio-environnementaux, la représentation générique de l'imperfection et enfin leur simulation en environnement dynamique en vue de prises de décision.

La première approche citée est la caractérisation d'un méta-modèle systémique intégrant des modèles complexes résultant de l'assemblage de modèles élémentaires via des fonctions, des flux et des interactions avec d'autres briques. Nous verrons que ce méta-modèle s'intègre de manière naturelle au langage standard de modélisation des systèmes SysML.

Nous verrons de manière détaillée dans l'état de l'art les approches analytiques ainsi que les approches systémiques actuelles utilisant ces derniers, au service du domaine des systèmes sociaux-environnementaux. Nous verrons que, pour la simulation de tels systèmes, la méthode de Monte-Carlo est souvent la seule approche utilisable pour l'étude des systèmes non linéaires de grande dimension pour lesquels aucune approche analytique n'est applicable (n simulations avec n très grand). Cependant, nous verrons que l'approche numérique comporte un certain nombre de limites. Nous étudierons également les apports potentiels du langage de modélisation standard SysML au domaine des systèmes socio-environnementaux et enfin nous soulignerons ses limites relatives à nos problématiques.

Le méta-modèle que nous décrivons utilise des modèles et des données issues de différentes sources et prend en compte les imperfections qui les entachent. Nous proposerons donc une approche à la fois opérationnelle et rigoureuse de représentation et de calcul sur des quantités dont l'imperfection ne relève pas seulement des approches probabilistes et statistiques, et qui s'intègre dans le méta-modèle précédent. Dans l'état de l'art, nous étudierons en quoi ces

représentations de l'imperfection nécessitent des traitements différenciés, mais qui doivent cependant pouvoir être utilisées conjointement, permettant ainsi de prendre en compte la réalité des situations. Par exemple, on distingue l'imprécision due au manque de données ou de connaissances de la variabilité naturelle d'un phénomène.

Enfin, notre approche consiste, à partir des informations hétérogènes et imparfaites, à bâtir un méta-modèle de décision. Ce méta-modèle permettra à un expert d'un domaine de proposer son propre modèle de décision grâce à des opérateurs et agrégats. Dans l'état de l'art, nous étudierons les bases de la théorie de la décision.

Nous verrons que les approches adoptées nécessitent un modèle architectural permettant la mise en œuvre d'applications basées sur les concepts que nous développerons en termes de modèles systémiques, avec la manipulation des données imparfaites.

Dans le paragraphe suivant, nous présenterons les contributions de notre thèse, et enfin les chapitres qui la constituent.



# Contributions

Les résultats des travaux de notre thèse s’articulent autour des trois contributions majeures suivantes :

- **Un méta-modèle systémique pour la modélisation, la simulation et l’ « optimisation durable » des systèmes socio-environnementaux.** Ce méta-modèle permet de décrire des éléments naturels dont la dynamique est plus ou moins bien connue ainsi que des éléments artificiels mis en œuvre par les acteurs du domaine destinés à exploiter ou contrôler les premiers. Nous proposons par ailleurs une correspondance naturelle entre les concepts proposés et les notations classiques du langage SysML<sup>4</sup>, tout en en apportant la rigueur sémantique indispensable.
- **Une approche formelle pour la représentation et la combinaison de données imparfaites.** Ceci comprend la représentation de données aléatoires et/ou imparfaitement connues, par le biais d’un formalisme unique, ainsi que la manière rigoureuse de les combiner par le biais de calculs algébriques. Nous montrons que ce formalisme s’intègre de manière naturelle dans le méta-modèle précédent et qu’il est également apte à s’intégrer dans l’extension du langage SysML définie dans le chapitre précédent.
- **Un modèle de décision** permettant à un expert d’un domaine de proposer son propre modèle de décision grâce à des opérateurs et agrégats adaptés aux modèles systémique et de représentation de l’imperfection précédemment développés.
- **Un modèle architectural** pour la validation, la mise en œuvre et l’opérationnalisation d’applications basées sur les concepts définis ci-dessus en termes de modèles de systèmes, avec manipulation de données imparfaites, dans un cadre de machines distribuées.

Tout au long de ce document, et plus particulièrement dans le chapitre sur l’opérationnalisation, nous illustrerons nos apports à travers l’exemple d’une exploitation

---

<sup>4</sup> Systems Modeling Language ; langage standard de modélisation graphique spécifique au domaine de l’ingénierie système

agricole type, pouvant aisément être extrapolé à tout système socio-environnemental. Ceci nous permettra de valider la pertinence des concepts proposés dans la présente thèse.

Dans la **première partie**, nous avons présenté le contexte, la problématique, l'approche et les contributions majeures autour desquels s'articule notre thèse.

La **deuxième partie** est organisée en trois chapitres. Elle dresse un état de l'art des approches existantes qui contribuent ou peuvent contribuer aux modélisations analytique et systémique dans le domaine des systèmes socio-environnementaux en contexte incertain. Le **premier chapitre** présente les bases théoriques et les applications utilisées pour la modélisation des systèmes socio-environnementaux, dans le cas où une solution analytique existe. Le **deuxième chapitre** est consacré à un état de l'art sur l'approche systémique, utilisée dans le cas où les lois sous-jacentes ne sont pas connues et avec des signaux d'entrée aléatoires, si bien que l'on doit réaliser des simulations (méthode de Monte Carlo). Nous aborderons aussi dans cette partie les apports de la modélisation via le langage standard SysML et leurs limites. Enfin, dans le **troisième chapitre**, nous ferons un bilan des approches d'aide à la décision utilisées dans le cadre des systèmes socio-environnementaux, avant de présenter et d'évaluer les approches mathématiques existantes de modélisation et manipulation d'informations imparfaites.

Cette deuxième partie est clôturée par un positionnement de notre approche par rapport aux travaux analysés, ce qui nous permettra de poser les principes fondateurs d'une approche systémique unifiée pour l'optimisation durable des systèmes socio-environnementaux.

La **troisième partie** est organisée en trois chapitres : le **quatrième chapitre** est dédié à la conceptualisation d'un méta-modèle systémique pour l'optimisation des systèmes socio-environnementaux, qui permet de modéliser, simuler et évaluer de tels systèmes à partir de « briques » élémentaires, et constituant un profil<sup>5</sup> SysML, le langage standard de modélisation des systèmes. Dans le **cinquième chapitre**, nous proposerons un formalisme unifié pour modéliser et manipuler les informations dont les différentes formes d'imperfection sont rencontrées dans le cadre des systèmes socio-environnementaux et s'intégrant dans le

---

<sup>5</sup> Un profil SysML est une extension utilisée pour personnaliser SysML pour un domaine spécifique par l'intermédiaire de mécanismes d'extension qui enrichissent la sémantique et la syntaxe du langage.

profil SysML précédemment réalisé. Enfin, dans le **sixième chapitre**, nous proposerons un modèle de décision en contexte incertain, i.e. différentes interprétations particulières de combinaisons de valeurs et de variateurs incertains ainsi que des opérateurs génériques d'agrégation sur la base du formalisme unifié que nous avons précédemment proposé, en vue de prendre des décisions.

Dans la **quatrième partie**, nous validerons nos contributions par la mise en œuvre de deux scénarios d'aide à la décision portant sur des exemples liés à l'agriculture dans le **septième chapitre**, puis nous opérationnaliserons notre approche dans le **huitième chapitre** grâce à une architecture orientée service que nous réaliserons sur la base des concepts précédemment proposés.

Enfin, une **conclusion** dresse une synthèse de nos travaux, de leurs apports potentiels au domaine des systèmes socio-environnementaux, ainsi que nos perspectives en termes de travaux futurs.

## Deuxième partie : état de l'art

# 1 Chapitre 1 : l'approche analytique : modélisation de fonctions de productions

Dans ce chapitre, nous présentons les bases théoriques et les applications utilisées pour la modélisation des systèmes socio-environnementaux. D'une manière générale, les scientifiques élaborent des modèles à partir de données brutes collectées, afin d'étudier et de prévoir le comportement de tels systèmes. Cette approche est valable dans le cas où une solution analytique existe.

## 1.1 Bases théoriques

Tout d'abord, l'approche analytique consiste à extraire des données et définir des macro-modèles, permettant de prédire le comportement d'un système socio-environnemental et de relier des variables entre elles. En effet, les scientifiques construisent des macro-modèles pour expliquer des phénomènes. Des relations, des causes et des explications liées aux observations de ces phénomènes sont recherchées. Les observations sont réalisées par rapport à l'évolution de certaines caractéristiques d'éléments qui peuvent être artificiels ou naturels : des grandeurs physiques (mesures, quantités de fertilisants, de semences, précipitations, temps) et des événements (météo, crises sanitaires).

Il s'agit d'une approche de type « Data Mining » ; extraction des données, analyse et interprétation par le scientifique. L'analyse de ces données consiste à utiliser la statistique, c'est-à-dire décrire un échantillon avec des agrégats, estimer des modèles de régression d'une variable expliquée sur une ou plusieurs variables explicatives, tester des hypothèses, et induire les caractéristiques inconnues d'une population à partir d'un échantillon issu de cette population, avec une certaine marge d'erreur. Cette exploitation de l'approche Data Mining [Liao et al., 2012] et du KDD<sup>6</sup> [Fayyad et al., 1996 (1)] [Fayyad et al., 1996 (2)] permet d'obtenir des macromodèles tels que des coefficients de linéarisation entre variables, la recherche de corrélations et de variables explicatives par des méthodes de régression. Les

---

<sup>6</sup> KDD est l'acronyme de Knowledge Discovery in Databases (extraction de connaissances à partir de données).

modèles obtenus ont pour but de prévoir l'état futur de variables qui caractérisent le système étudié. Le KDD est le processus informatique global de découverte de connaissances à partir de grandes quantités de données, allant de l'extraction des données brutes, leur prétraitement, jusqu'à leur interprétation par le scientifique. Ce processus est décrit dans l'illustration ci-dessous :

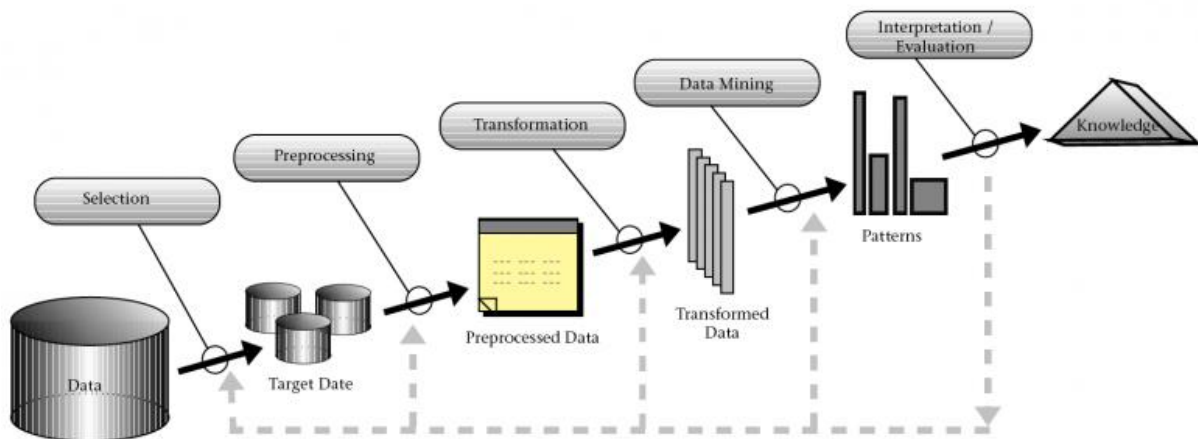


Figure 1 : vue d'ensemble des étapes du processus de KDD (source : [Fayyad, U. et al, 1996])

Le processus de KDD comprend les étapes suivantes : sélection, prétraitement, transformation, fouille et interprétation par l'expert. L'étape clé de ce processus est le Data Mining ou fouille de données, processus mathématique de découverte d'information à partir de grandes quantités de données, utilisé pour résoudre des problèmes dans de nombreux domaines tels que le marketing (étude des attentes et de la satisfaction des clients, prédiction du comportement des clients), Internet (moteurs de recherche, publicités ciblées), la banque et la finance (analyse des risques, des clients), la médecine et l'assurance maladie (études épidémiologiques, détection de fraudes à l'assurance), etc. Ces techniques revêtent une importance majeure à l'heure où le nombre et la taille des bases de données augmentent de manière exponentielle.

Les chercheurs établissent alors des modèles de fonctions de production grâce à différentes méthodes mathématiques d'analyse de données. Dans les méthodes supervisées, le système apprend à classer les éléments selon un modèle de classement qui contient des groupes prédéfinis, au contraire des méthodes non supervisées où les classes ne sont pas prédéfinies (clustering). Il existe un certain nombre de méthodes supervisées, à savoir (liste non exhaustive) : les méthodes de régression, l'analyse discriminante, les arbres de décision, les forêts aléatoires, les réseaux de neurones supervisés, etc.

De ces méthodes, on pourra déduire un modèle qui détermine les sorties du système en fonction des variables d'entrées. Un modèle comporte les éléments suivants : une fonction, des variables d'entrée et des paramètres. Ces fonctions de production sont en général des approximations locales de la réalité liées à des zones géographiques spécifiques, telles que des modèles de rendement et de productivité en fonction de paramètres explicatifs tels que la quantité de semences, d'engrais ou de pluie. Voici la forme générale d'un modèle :

$$Y = f(X, \theta) + \varepsilon$$

Avec :

- $Y$  : variable à expliquer, calculée par le modèle
- $f$  : fonction de production
- $X$  : variables d'entrée mesurées ou observées.
- $\theta$  : paramètres : valeurs numériques qui ne sont ni des variables d'entrée ni calculées par le modèle. Leurs valeurs sont estimées à partir de données expérimentales et/ou d'avis d'experts.
- $\varepsilon$  : erreur dont l'espérance est nulle par hypothèse et de même variance que  $Y$ . Lors de l'estimation des paramètres, on cherche à minimiser sa variance.

La performance d'un modèle dépendra (1) de la quantité de données, (2) de leur qualité et (3) de la méthode utilisée pour estimer ses paramètres.

## 1.2 Applications aux systèmes socio-environnementaux

Dans ce paragraphe, nous présentons les méthodes de modélisation par estimation de paramètres généralement appliquées au domaine des systèmes socio-environnementaux. Tout d'abord, nous présenterons un exemple très simple de modélisation de la quantité d'azote absorbée par le colza en fonction de la dose d'engrais appliquée [Makowski, 2005]. On suppose que ce modèle inclut une variable d'entrée continue «  $x$  » (qui correspond à la dose d'engrais appliquée) et deux paramètres  $\theta_1$  et  $\theta_2$  inconnus.

Dans ce modèle, la dose d'engrais agit linéairement sur le rendement :

$$y = \theta_1 + \theta_2 \cdot x$$

On dispose d'une série de mesures d'azote absorbé obtenues sur un certain nombre de parcelles de colza de la population d'intérêt (par exemple une région agricole).

Pour estimer les paramètres du modèle, on utilise une méthode de régression linéaire (ici la méthode des moindres carrés ordinaires). Les estimateurs des paramètres sont les valeurs de  $\theta_1$  et  $\theta_2$  qui minimisent l'expression suivante :

$$\sum_{i=1}^n (y_i - \theta_1 - \theta_2 \cdot x_i)^2$$

On évalue la précision des estimateurs (1) en estimant la variance des estimateurs à partir des données et (2) en analysant les résidus<sup>7</sup> afin de vérifier leur indépendance et l'homogénéité de leurs variances. Si ces dernières ne sont pas vérifiées, l'estimateur des moindres carrés ordinaires n'est pas la méthode permettant de trouver des estimateurs de variance minimale (théorème de Gauss-Markov). Il faudra alors sélectionner les paramètres à estimer ou bien utiliser d'autres méthodes.

De nombreux autres exemples de fonctions de production ont été réalisés par les experts du domaine des systèmes socio-environnementaux [Magnan de Bornier, 2015] [Dury, 2011] [Araujo et al., 1999]. Ces fonctions de production sont en général des approximations locales de la réalité liées à des zones géographiques spécifiques, telles que des modèles de rendement et de productivité en fonction de paramètres explicatifs tels que la quantité de semences, d'engrais ou de pluie : *Productivité* =  $F(QSemences, QEngrais, QPluviométrie)$ , dont une approximation (linéarisation) est *Productivité* =  $K_1 \cdot QSemences + K_2 \cdot QEngrais + K_3 \cdot QPluviométrie + K_4$  pour des intervalles [QS1, QS2], [QP1, QP2] et [CC1, CC2] donnés, mais aussi des modèles et fonctions de production et de profits à maximiser en fonction des applications d'engrais NPK tels que : *Production* =  $0.4 * X_1 + 0.5 * X_2 + 0.2 * X_3 - 0.002 * X_1^2 - 0.004 * X_3^2 - 0.001 * X_1 * X_2 - 0.001 * X_1 * X_3 - 0.0004 * X_2 * X_3 + 23.5$  et *Profit* =  $5000y - 150(X_1 + X_2 + X_3) - 150000$  avec  $X_1, X_2, X_3$  étant respectivement les applications d'engrais N, P et K ; 5000 le cours des productions en €/kg ; y la production en kg ; 150 le prix des engrais NPK et enfin 150000 les coûts fixes (en €) [Wolff, 1996].

---

<sup>7</sup> Un résidu est un écart entre une observation et la valeur correspondante prédite par le modèle.



### 1.3 Limites de l'approche analytique

Dans certains cas, il n'existe pas de relation analytique entre estimateurs et données permettant de modéliser les phénomènes observés. En effet, les structures en jeu peuvent être complexes, avec de nombreuses variables et paramètres en jeu (avec une quantité souvent insuffisante de données), qui peuvent être corrélées, et des modèles non linéaires. Ce type de problème est très complexe et nécessite un modèle dynamique. Dans ce cas, un modèle de réponse de type  $y = f(x_1, \dots, x_n, \theta_1, \dots, \theta_n) + \varepsilon$  nécessite de sélectionner les paramètres à estimer mathématiquement, les autres étant évalués par exemple à partir d'une revue de la littérature. Les inconvénients de cette approche sont que la sélection des paramètres à estimer d'une manière ou d'une autre est subjective (elle dépend uniquement de l'avis de l'expert à un instant donné) et qu'il est difficile de trouver une concordance satisfaisante entre les situations considérées dans la littérature et celles qui concernent la situation à modéliser [Makowski, 2005]. L'autre forme de modèle dynamique est la modélisation sous forme de système dynamique, de type  $y_{n+1} = y_n + f(x_1(n-1), \dots, x_n(n-1), \theta_1, \dots, \theta_n) + \varepsilon$ . Dans ce cas, il est possible d'estimer les paramètres en faisant un grand nombre de simulations, classiquement par la méthode de Monte Carlo. Dans la suite de cet état de l'art, nous étudierons l'approche systémique et ses limites dans le cadre des systèmes socio-environnementaux.

## 2 Chapitre 2 : l'approche systémique : modélisation et simulation de systèmes

Il convient en premier lieu de définir la notion de système. D'une manière générale, un système est un ensemble d'éléments en interaction dynamique organisés en fonction d'un but [De Rosnay, 1975]. Les bases de l'approche systémique ont été posées par Jean-Louis Le Moigne [Le Moigne, 1977]. Chaque système possède un certain nombre de caractéristiques qui lui sont propres : son ouverture correspond à l'ensemble de ses interactions permanentes avec son environnement qu'il modifie et qui le modifie en retour, souvent de manière aléatoire. Un système est constitué des éléments qui le composent, qui sont eux-mêmes en interaction permanente. Classiquement, on distingue l'aspect structurel (i.e. l'organisation dans l'espace des éléments, les stocks, etc.) de l'aspect fonctionnel (caractérisation dynamique des éléments dépendant du temps, i.e. les flux d'énergie ou de matière, les échanges, les délais, etc.). De plus, un système est en général décomposable en sous-systèmes constitutifs et est finalisé, i.e. que le comportement et l'équilibre d'un système sont maintenus par des boucles de rétroaction.

Dans ce chapitre, nous présentons les bases de l'approche systémique notamment en termes de buts, de fonctions, de performance, de structure, etc. Dans ce cas, les entrées du système peuvent être incertaines et/ou en quantité modeste. De plus, les lois sous-jacentes modélisant le comportement du système ne sont pas forcément connues : il n'y a pas de loi ni de distribution connue concernant le système à modéliser. La solution est alors de réaliser un grand nombre de simulations avec tirages aléatoires des signaux d'entrée, en observant les sorties obtenues en fonction de ces entrées afin d'estimer les paramètres du système simulé. Nous étudierons ensuite la modélisation de systèmes via le langage standard de modélisation des systèmes (SysML) ainsi que ses apports potentiels pour notre sujet d'étude. Nous donnerons ensuite un aperçu des méthodes de simulation de systèmes. Nous concluons enfin ce chapitre en mettant en lumière les limites des approches étudiées.

## 2.1 Les modèles systémiques existants pour la simulation de systèmes socio-environnementaux

Les modèles systémiques dédiés aux systèmes socio-environnementaux permettent de simuler le comportement réel du système considéré afin d'établir ses états futurs possibles. Ses réponses possibles sont fournies en fonction de variables d'entrées à la fois nombreuses, interdépendantes et ayant de grandes variabilités, le système réel évoluant continuellement.

Plus spécifiquement, dans le domaine de l'agriculture, il existe de nombreux outils dédiés à la simulation et/ou à l'évaluation d'exploitations agricoles, qui se distinguent entre eux par leur application à des contextes d'utilisation différents : le type de production (cultures, animaux, maraîchage, etc.), l'échelle d'analyse (exploitation, parcelle, etc.), le coût et la durée de mise en œuvre (quelques heures à plusieurs jours), le public visé (agriculteur, technicien, conseiller, chercheur, etc.). Au centre de l'exploitation agricole se situe l'agrosystème, qui correspond à l'ensemble « sol – culture – atmosphère (dont les organismes vivants) ». Il comporte de nombreuses interactions de natures chimiques (e.g. le cycle de l'azote, les transferts sol/plante), physiques (e.g. la gestion de l'eau, la structure et le travail du sol), et biologiques (e.g. les bioagresseurs, la fertilisation, les organismes vivants). L'agrosystème est en évolution permanente de par les effets de facteurs externes (e.g. le climat) ou internes (e.g. les techniques agronomiques appliquées). L'agrosystème est également lié à des enjeux économiques, environnementaux et sociaux. Afin d'évaluer la globalité du système, on utilise des outils d'évaluation multicritères. L'agent qui agit sur ce système complexe de par ses décisions est l'agriculteur exploitant, d'où la nécessité d'outils d'aide à la décision ainsi que de simulations permettant d'évaluer les réponses possibles en sortie du système en fonction des entrées sur lesquelles l'agriculteur a agi. En effet, un des buts principaux de l'agriculteur est de trouver le système qui optimise son revenu, qui est fonction de nombreux facteurs comme les prix du marché, les coûts, la quantité, la production, etc. De plus, l'agrosystème est amené à évoluer : les variables d'entrée sur lesquelles l'agriculteur peut agir peuvent être ajustées en fonction des variables de sortie voulues. Enfin, la quasi-totalité des variables sont interdépendantes. En effet, le revenu peut être modifié en fonction de nombreux facteurs comme les prix du marché, les coûts, la quantité, la production, etc. Il existe de nombreuses modélisations des concepts d'agrosystème [Meynard et al., 2001] [Doré et al., 2006]. Ceux-ci sont souvent représentés sous forme de schémas ne représentant qu'un seul aspect du système

(en général l'aspect fonctionnel), sans standard de représentation dans un objectif d'automatisation et de portabilité.

Certains modèles ont été implémentés sous forme de logiciel. Parmi les outils existants, Olympe [Attonaty, 2010] [Penot et Deheuvels, 2007] est à la fois un simulateur technico-économique et un outil d'aide à la décision fournissant des indicateurs économiques tels que le bilan, la trésorerie, des graphiques et états de sortie paramétrés par l'utilisateur, à partir de données d'entrée telles que la production, des règles de calcul, des incertitudes sur les prix, certaines caractéristiques du cheptel, l'assolement, etc. Il est donc possible de simuler plusieurs cas possibles en fixant des paramètres d'entrée (météo, modification du système de production ou de la réglementation) afin de tester des hypothèses et d'en mesurer les conséquences économiques. Cet outil spécifique ne permet ainsi de simuler que l'aspect économique de l'exploitation agricole.

L'outil STICS<sup>8</sup> [Brisson et al., 1998] [Brisson et al., 2002] [Brisson et al., 2003] est un simulateur d'agrosystèmes au cours d'une année, à pas de temps journalier. Il est développé, depuis 1996, à l'unité de bioclimatologie de l'INRA d'Avignon, avec la participation d'autres instituts tels que l'unité d'agronomie de l'INRA de Laon, le CIRAD, le CEMAGREF, etc. La plupart des modèles agronomiques simulés sont issus de modèles existants. Il est adaptable à de nombreux types de cultures (blé, maïs, soja, sorgho, lin, prairies, tomate, betterave, tournesol, pois, colza, fraisier, carotte, laitue, etc.). Les paramètres d'entrée décrivent l'atmosphère (climat, rayonnement solaire, température, pluie, évapotranspiration de référence, vent et humidité), le sol (considéré comme une succession de couches caractérisées par leur réserve en eau, en azote minéral et en azote organique), les cultures (biomasse aérienne, teneur en azote, indice foliaire, nombre, biomasse et teneur en azote des organes récoltés, les pratiques culturales et leurs interactions (densité racinaire dans le sol, etc.). Les paramètres de sortie sont relatifs à la production (en termes de quantité et de qualité), à l'environnement et à l'évolution des caractéristiques du sol sous l'effet de la culture. Les principaux processus simulés sont la croissance et le développement des cultures ainsi que les bilans hydrique et azoté. La croissance de la culture est pilotée par le bilan de carbone : interception du rayonnement solaire par le feuillage puis transformation en biomasse aérienne qui est orientée vers les organes de récolte pendant la phase finale du cycle. Le bilan azoté de

---

<sup>8</sup> Simulateur multidisciplinaire pour les cultures standard

la culture est simulé parallèlement au bilan de carbone, dont il dépend en partie. Selon le type de plante, le développement de la culture est piloté soit par un indice thermique (en degrés-jours), soit par un indice photothermique (exigences photopériodiques) ou vernalo-photothermique (besoins en froid de la plante dans le cas du blé). Le modèle de développement est utilisé pour (1) faire évoluer l'indice foliaire et (2) définir la phase de remplissage des organes récoltés. L'existence éventuelle de stress (déficits hydrique ou azoté, températures gélives ou échaudantes, excès d'eau) est prise en compte au travers d'indices qui peuvent réduire la croissance végétative et celle des organes de stockage. Ces indices sont calculés dans les bilans hydrique, azoté ou énergétique. L'effet des techniques culturales (itinéraires techniques appliqués) sur le fonctionnement du système sol-culture-atmosphère doit également être pris en compte.

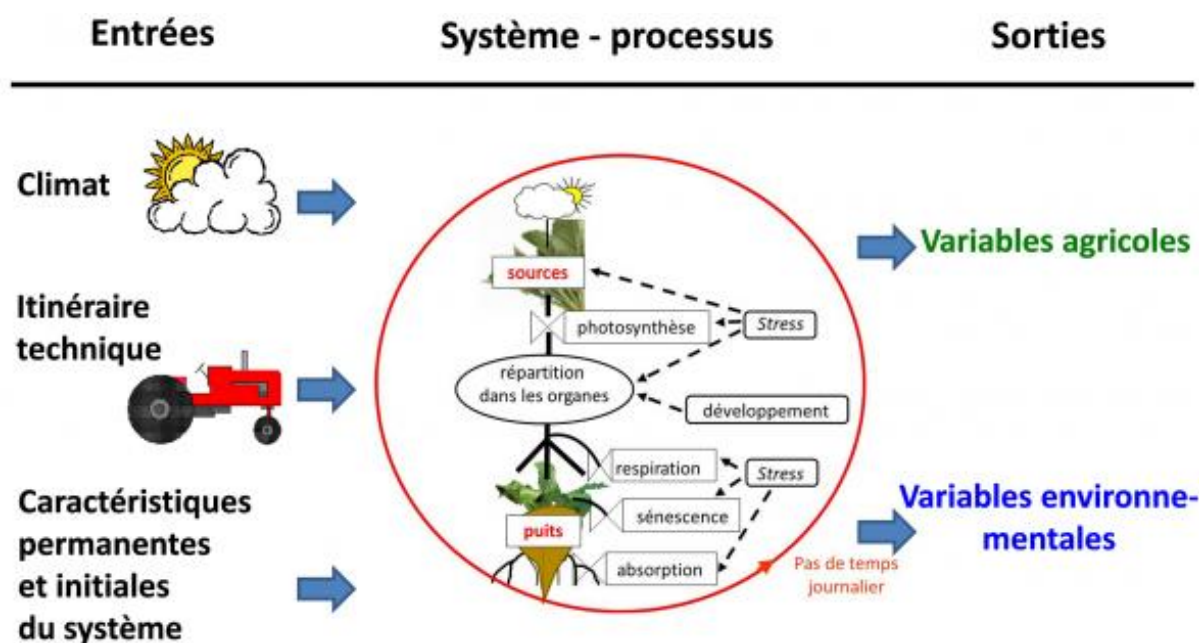


Figure 2 : principe de fonctionnement de STICS<sup>9</sup>

Ces modèles systémiques comportent plusieurs limites. Tout d'abord, ils ont l'inconvénient majeur d'être d'autant plus complexes que le modèle se rapproche de la réalité. Ainsi, la difficulté de modélisation peut dépasser l'enjeu. Il n'y aurait en conséquence aucun intérêt pratique à passer plusieurs années de projet pour modéliser, simuler et optimiser une unique exploitation agricole. De plus, de tels modèles manquent de souplesse puisqu'ils sont en

<sup>9</sup> Source : <http://www6.paca.inra.fr/stics/Qui-sommes-nous/Presentation-du-modele-Stics/Les-principes-de-fonctionnement> (page consultée le 01/03/2016).

général spécifiques à un domaine précis tout en s'appuyant en général sur des modèles de données statiques comme par exemple des modèles relationnels [Dantan et al., 2012 (2)]. L'ajout ou la modification de variables ou d'objets simulés peuvent alors entraîner des profondes modifications dans les résultats des modèles simulés, qui ne peuvent être réalisées que par des experts en informatique et du métier. Ce manque de souplesse peut également être souligné par l'aspect « boîte noire » des différents modèles : les modèles simulés, complexes, sont en général imposés à l'expert du domaine. Il faudrait donc modéliser les différents aspects du même système avec un langage unifié de modélisation des systèmes, ce que nous allons présenter dans le prochain paragraphe.

## 2.2 L'approche systémique

L'approche systémique est une méthodologie de représentation, de modélisation d'un ensemble d'éléments actifs en interaction dynamique avec son environnement par l'intermédiaire de flux. Cette approche permet aux experts de différents domaines d'améliorer leurs connaissances sur des objets d'études en synthétisant leurs principales caractéristiques et en simulant leurs actions futures. Les flux peuvent être (1) connus à tout instant ; il s'agit alors de signaux à temps continu pouvant être modélisés mathématiquement par des fonctions du temps  $x(t)$  ou bien (2) échantillonnés. Dans ce dernier cas, il s'agit de signaux à temps discret connus seulement à certains instants, pouvant être modélisés par une suite  $x(n)$ . Une telle modélisation systémique répond à une finalité ; elle peut avoir ainsi pour but de décrire un existant ou bien de répondre à une certaine préoccupation de conception concernant le futur système. Dans ce cadre, les aspects possibles d'un système [Pollet, 2015] incluent en particulier :

- Les aspects fonctionnel et comportemental, qui consistent à décrire ce que doit faire le futur système et comment il devra réagir à des sollicitations de l'extérieur ;
- Les aspects liés aux performances. Des modèles doivent alors rendre compte du fonctionnement de manière quantitative ;
- Les aspects structurels, qui consistent à décrire de quoi sera fait physiquement le système, avec les interactions entre les constituants ainsi mis en évidence ;
- D'autres aspects, tels que les coûts de revient, ou encore la fiabilité, etc. dont des modèles spécifiques doivent être capables de rendre compte.

La figure ci-après illustre un exemple de système de production agricole (agrosystème) :

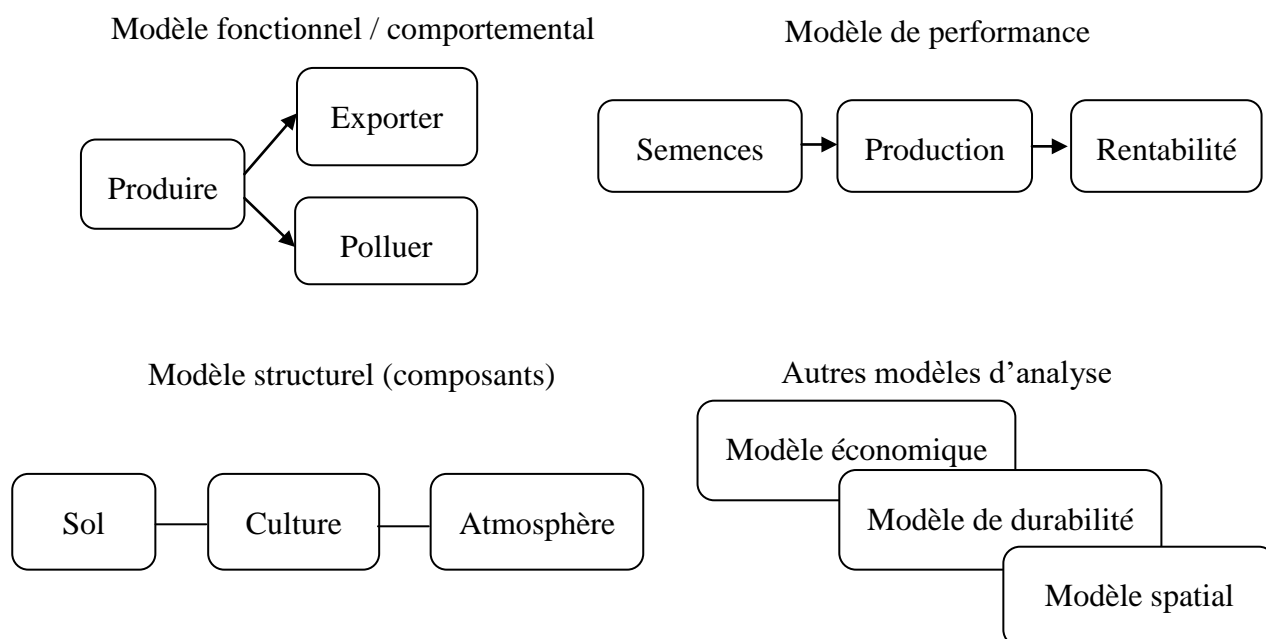


Figure 3 : les différents aspects de la modélisation d'un système socio-environnemental

La modélisation dite intégrée consiste à formaliser les différents aspects du même système grâce à des modèles, réalisés avec un langage unifié de modélisation des systèmes, sur la base d'une cohérence entre ces différents aspects. La pluralité des applications possibles nécessite d'utiliser des descriptions communes et non « propriétaires » des différents domaines en jeu. Enfin, le rôle central de la simulation dans le domaine de la systémique nécessite que ces descriptions puissent mener de manière aisée à des modèles directement exécutables par des logiciels.

## 2.3 SysML, le langage standard de modélisation des systèmes

### 2.3.1 Vue générale

Dans ce paragraphe, nous ferons une brève introduction au langage standard de modélisation des systèmes, c'est-à-dire SysML, puis nous étudierons ses apports potentiels pour notre sujet d'étude, à savoir l'étude des systèmes socio-environnementaux.

SysML<sup>10</sup> est le langage standard de modélisation des systèmes. Il est basé sur le langage UML version 2, dont il reprend plusieurs diagrammes, tout en en proposant de nouveaux ou en

---

<sup>10</sup> Systems Modeling Language, ou « langage pour la modélisation des systèmes ». Lien vers le site Web officiel de SysML : <http://www.sysml.org> (page consultée le 01/03/2016).

adaptant d'autres. Cette caractéristique fait de ce langage un profil UML. Il est spécifiquement adapté à la modélisation des systèmes (et non plus seulement des logiciels). Il est issu d'une initiative conjointe par l'INCOSE et l'OMG<sup>11</sup>. La version actuelle (1.2) date de juin 2010. Il existe de nombreuses implémentations d'éditeurs SysML disponibles. Ce langage, s'il n'est pas encore aussi répandu que le langage UML, est maintenant le langage standard de modélisation pour l'ingénierie système, et est donc de plus en plus utilisé.

Le langage SysML supporte la spécification, l'analyse, la conception, la vérification, et la validation de systèmes incluant du matériel, du logiciel, des données, des humains, des procédures, et des services extérieurs (facilities). Précisons que SysML est un langage de modélisation visuel ; il prescrit : (1) une sémantique, c'est-à-dire un ensemble de concepts et de significations et (2) des notations, qui sont les représentations visuelles de ces significations. Cependant, SysML est indépendant de toute méthodologie ou outil.

Pour les besoins de l'ingénierie système, SysML a retenu tels quels certains diagrammes d'UML (par exemple le diagramme d'états). D'autres diagrammes ont été enrichis (par exemple le diagramme d'activités) ou sont complètement nouveaux (par exemple le diagramme d'exigences).

---

<sup>11</sup> Object Management Group (site Web de l'OMG pour le langage SysML : <http://www.omgsysml.org>). SysML - tout comme UML - est décrit par un métamodèle conforme au MOF (Meta-Object Facility, <http://www.omg.org/mof/>), qui est un standard de l'OMG (pages consultées le 01/03/2016).



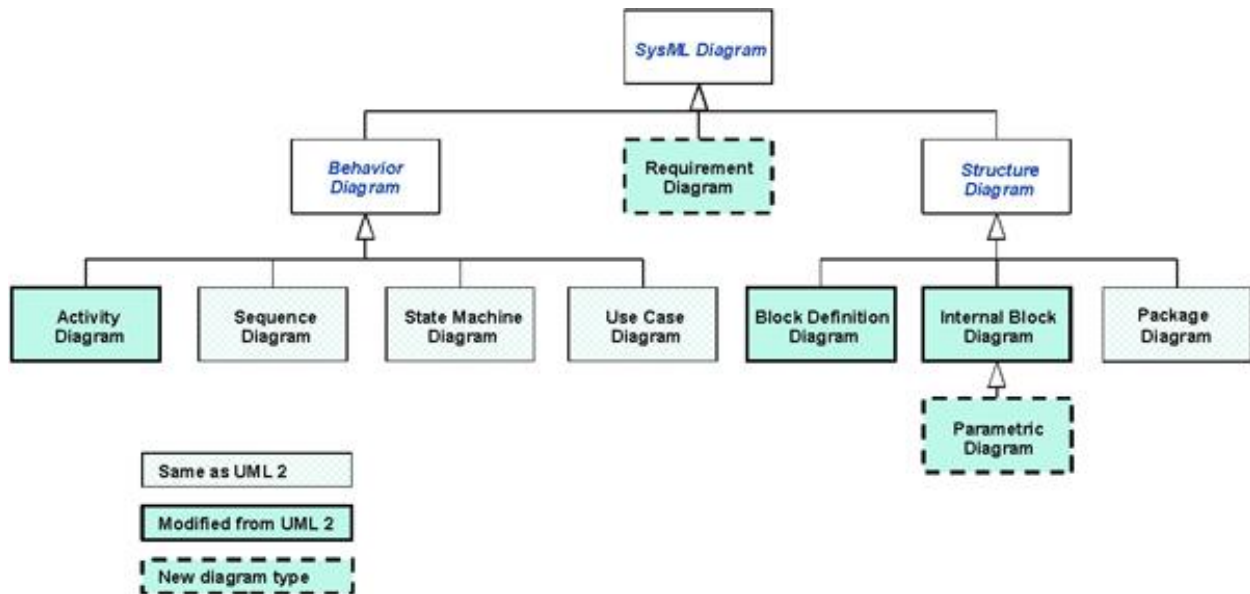


Figure 4 : les différents diagrammes SysML (source : [OMG, 2015])

Cet ensemble peut être divisé en trois grands groupes :

- Quatre diagrammes dynamiques et comportementaux (voir tableau ci-dessous).
- Quatre diagrammes structurels (voir tableau ci-dessous).
- Un diagramme transverse (c'est-à-dire n'entrant pas dans l'une des deux catégories ci-dessus) : le diagramme d'exigences, qui représente les exigences applicables à un système ou à un constituant, ainsi que les relations entre celles-ci. Il s'agit d'un nouveau diagramme.

Tableau 1 : diagrammes dynamiques et comportementaux de SysML

Diagramme	Lien avec UML 2	Description
<b>Diagramme d'activités</b>	Repris d'UML 2 avec des extensions	Description sous forme de flux ou d'enchaînement d'activités du comportement du système.
<b>Diagramme de séquence</b>	Identique au diagramme de même nom UML 2.	Représentation du déroulement des traitements et des interactions entre des éléments.
<b>Diagramme d'états</b>	Identique au diagramme de même nom UML 2	Description sous forme de machine à états finis du comportement du système ou d'un composant.

<b>Diagramme de cas d'utilisation</b>	Identique au diagramme de même nom UML 2.	Identification des points d'interaction entre les acteurs et le système et les fonctionnalités que doit fournir celui-ci.
---------------------------------------	---	---

Tableau 2 : diagrammes structurels de SysML

Diagramme	Lien avec UML 2	Description
<b>Diagramme de définition de bloc</b>	Diagramme nouveau dans son usage, mais basé sur le diagramme de structure composite <sup>12</sup> .	Description des briques de base du système: blocs, compositions, associations, attributs, opérations, généralisation, etc.
<b>Diagramme de bloc interne</b>	Diagramme nouveau dans son usage, mais basé sur le diagramme de structure composite.	Description de l'organisation interne d'un élément.
<b>Diagramme de paquetages (ou package)</b>	Identique au diagramme de même nom UML 2.	Conteneur logique permettant de regrouper et d'organiser les éléments dans un modèle.
<b>Diagramme paramétrique<sup>13</sup></b>	Nouveau diagramme.	Représentation des contraintes d'un système et/ou des équations qui les régissent.

En résumé, quatre diagrammes ont été repris tels quels d'UML 2, trois ont été étendus et deux sont complètement nouveaux.

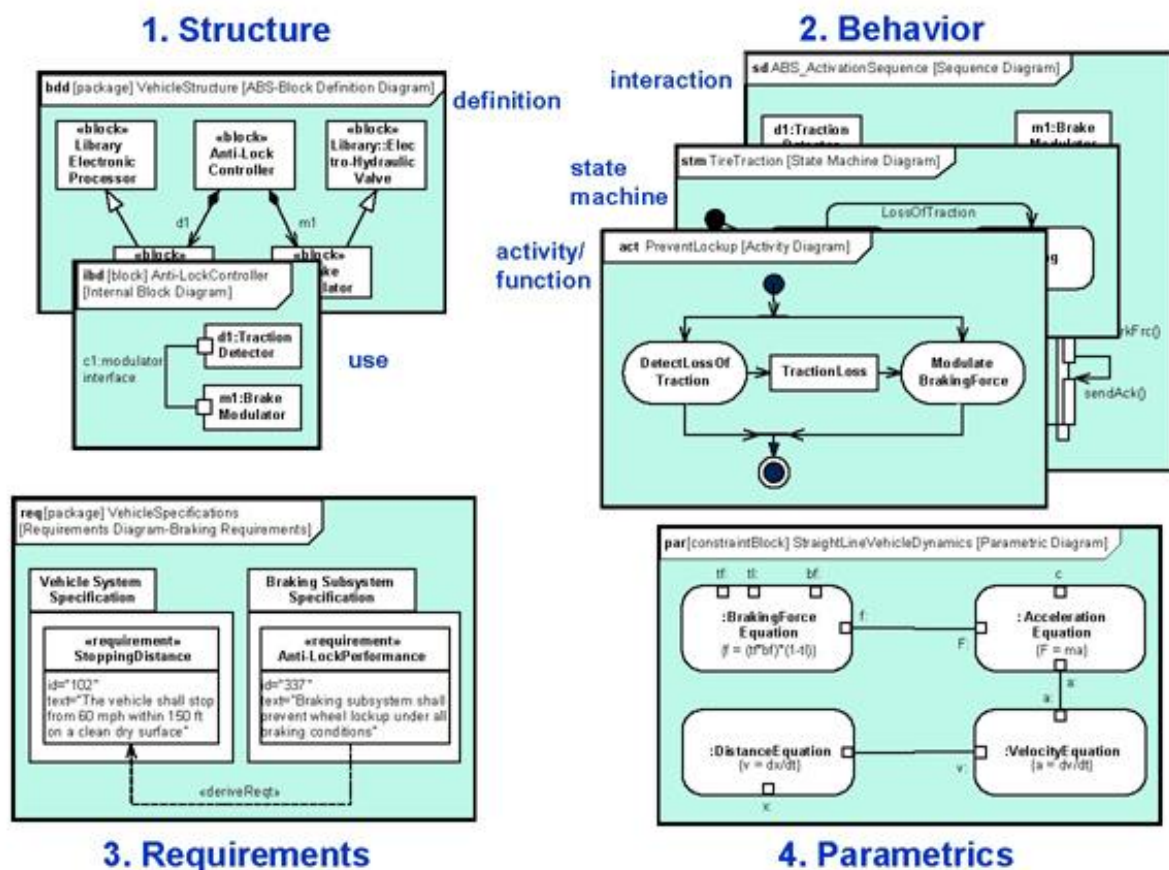
---

<sup>12</sup> Le diagramme de structure composite permet, si nécessaire, de décrire les relations entre les composants internes d'une classe donnée (vision « boîte blanche » de cette classe).

<sup>13</sup> Le diagramme paramétrique est classé par les promoteurs du langage comme un diagramme de structure. Mais il pourrait sembler tout aussi naturel de le classer comme un diagramme transverse. En effet, il ne représente pas vraiment la structure d'un système.

Sur le plan des concepts, le langage SysML repose sur quatre ensembles fondamentaux, qui constituent les quatre piliers de SysML, illustrés par la figure ci-dessous :

- La notion de structure. Pour représenter la structure, on utilise la notion de bloc, qui permet de modéliser tout élément concret : système, sous systèmes, composants logiciels ou matériels, mais aussi acteurs, système externes, ou encore éléments de matière, d'énergie ou d'information échangés. Les diagrammes correspondants sont les diagrammes de définition de bloc et de bloc interne ;
- La notion de comportement, pour la modélisation dynamique ou comportementale du système ou d'un constituant. Les diagrammes correspondants sont les diagrammes d'activités et d'états (et au besoin de séquence) ;
- La notion d'exigence. Le diagramme correspondant est le diagramme d'exigences ;
- Les équations et contraintes régissant le système ou ses constituants. Le diagramme correspondant est le diagramme paramétrique.



Note that the Package and Use Case diagrams are not shown in this example, but are respectively part of the structure and behavior pillars

Figure 5 : les quatre piliers de SysML (source : [OMG, 2015])

Un diagramme SysML représente un élément de modèle. Tout diagramme SysML sera contenu dans un cadre (Diagram Frame), comportant une entête (header) et un contenu (contents). Le contexte du diagramme est indiqué dans l'entête (« header »). Il comporte les informations suivantes :

- Un code indiquant le type de diagramme (Diagram kind) (act, bdd, ibd, sd, etc.). Il s'agit des différents types de diagrammes déjà indiqués, et que l'on va décrire dans la suite (« act » pour « Activities », « bdd » pour « Block definition diagram », etc.) ;
- Le type d'élément modélisé (Model element type) (package, block, activity, etc.) ;
- Le nom de l'élément modélisé ;
- Le nom du diagramme (défini par l'utilisateur qui conçoit le modèle).

Nous insisterons plus particulièrement sur les notions que nous allons aborder dans les contributions de notre thèse, c'est-à-dire : les diagrammes des blocs ainsi que les diagrammes paramétriques, ces deux derniers étant issus du standard SysML.

### 2.3.2 Modélisation de la structure d'un système socio-environnemental

En SysML, le bloc (« block ») constitue l'élément de base pour la modélisation de la structure d'un système. Un bloc peut représenter un système complet, un sous-système ou encore un composant élémentaire, logiciel ou matériel, voire un élément échangé (information ou élément physique). Les blocs sont eux-mêmes décomposables et peuvent posséder un comportement. Le diagramme de définition de bloc permet de faire apparaître les propriétés externes d'un bloc ainsi que les parties qui le constituent, sans toutefois entrer dans la structure des relations entre ces mêmes parties. Le diagramme de définition de bloc permet de décrire la hiérarchisation du système en constituants. Les propriétés sont les caractéristiques structurelles de base des blocs. Elles peuvent être de deux types principaux :

(1) Les valeurs (« value properties »), qui sont utilisées pour modéliser les caractéristiques quantitatives des blocs en terme de types de valeurs (« value types »). Elles sont basées sur les types de base proposés par SysML, à savoir : les types primitifs, les énumérations et les types structurés ; elles sont aussi quantifiables avec une distribution de probabilité, avec la syntaxe suivante :

```
<<distribution_name>>{parameters} {property name}: {unit name}
```

<<distribution\_name>> représente la loi de probabilité (<<uniform>>, <<normal>>) ou simplement un intervalle (<<interval>>).

(2) Les parties (« part properties »), qui décrivent la hiérarchie de décomposition du bloc en termes de blocs constitutants (ou relation « tout-partie »), dans laquelle un bloc représente le tout, et les autres blocs ses parties. Une instance du tout peut contenir plusieurs instances d'une partie, grâce à la notion de multiplicité.

Par exemple, si l'on considère le bloc « Parcelle », on définit sa valeur de surface. Nous pouvons typer cette propriété par un « value type » « surface », défini par une distribution de probabilité à valeurs réelles et une unité « hectare » ; par exemple :

```
<<uniform>>{min=10,max=12} surface:ha
```

Le bloc exploitation agricole est obligatoirement décomposable en blocs « parcelle ». De même, pour le bloc parcelle, on peut définir la partie sol. Chaque bloc (ou type de bloc) définit un ensemble d'instances partageant les propriétés du bloc, mais possédant chacune une identité unique.

Dans le bloc « exploitation agricole », l'exploitation agricole correspond au tout ; les parcelles aux parties (relations de composition).

On définit également des relations d'agrégation entre, d'une part, le bloc « parcelle » et, d'autre part, les animaux (par exemple le bétail) et les végétaux qu'il peut éventuellement contenir. Cette relation est moins forte que la relation de composition (composante facultative) et une même instance peut faire partie de plusieurs autres blocs (et non d'un seul). Dans ce cas là, du côté du tout, la partie est appelée « propriété partagée », ou référence.

On définit ensuite des relations de généralisation entre, d'une part, le bloc bétail, et, d'autre part, le bloc bovin et le bloc porcin qui « spécialisent » ce bloc. Ils « héritent » des propriétés du bloc généralisé et peuvent comporter des propriétés spécifiques supplémentaires. En SysML, la généralisation se représente graphiquement par une flèche triangulaire pointant sur le bloc généralisé, comme la spécialisation (héritage) de classes UML.

On définit aussi une relation d'association entre les blocs exploitation agricole et personne pour indiquer qu'une personne peut posséder une exploitation agricole. Il s'agit d'une relation d'égal à égal.

Un bloc peut posséder des propriétés comportementales. Les principales sont appelées « opérations » et peuvent être soit synchrones (l'émetteur est bloqué en attente d'une réponse) avec si besoin des paramètres (arguments) en entrée, en sortie, ou entrée-sortie, soit asynchrones (l'émetteur n'est pas bloqué en attente d'une réponse) : chaque réception est

associée à un signal qui définit un message entrant avec d'éventuels paramètres. Les signaux peuvent être définis comme des blocs avec le mot-clé «signal».

Les ports font partie de la définition des blocs. Il est donc possible de les faire apparaître dans le diagramme de définition de bloc. Les ports définissent les points d'interaction fournis (« provided ») ou requis (« required ») par un bloc. Un bloc peut avoir plusieurs ports qui spécifient des points d'interaction différents. Ils peuvent être représentés dans le diagramme de définition de bloc directement dans un compartiment spécifique (<<Flowport>>).

Comme nous l'avons précisé ci-dessus, le diagramme de définition de bloc ne met pas en évidence la structure des relations entre les parties qui le constituent. Pour palier cette limite, SysML définit le diagramme interne de bloc qui permet de modéliser les relations entre blocs.

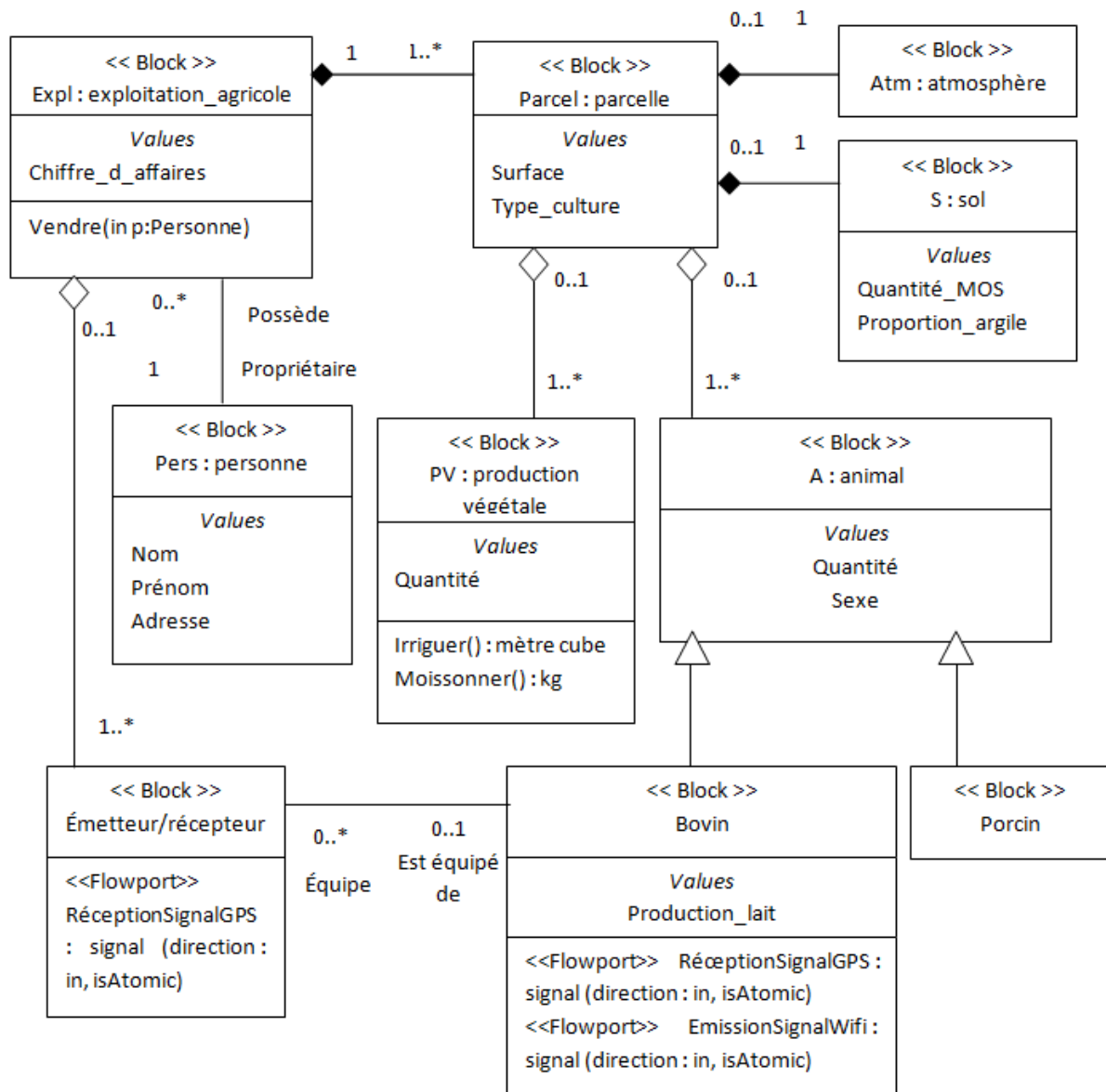


Figure 6 : exemple de diagramme de définition de bloc modélisant un agrosystème

### 2.3.3 Modélisation de la structure des relations entre les éléments d'un système

Le diagramme interne de bloc décrit les structures internes des blocs en termes de parties, ports et connecteurs. Il spécifie l'interconnexion des parties qui composent un bloc. L'intérêt du diagramme de bloc interne est de pouvoir relier les ports des différents blocs au moyen de connecteurs, ce que ne permet pas le diagramme de définition de bloc, qui définit uniquement des ports sans les connecter.

On peut représenter la connexion entre les éléments (« parts ») d'un bloc au moyen d'un diagramme de bloc interne. Un tel diagramme montre principalement les relations entre les

éléments de même niveau, ainsi que les éventuelles multiplicités des parties. Le cadre représente le bloc englobant. Il fournit le contexte pour tous les éléments du diagramme. Chaque extrémité d'une relation de composition, ou élément du compartiment « parts » apparaît comme un bloc à l'intérieur du cadre. Le nom du bloc est identique à celui qui apparaît dans le compartiment « parts ». On définit ainsi l'exploitation agricole comme bloc englobant, puis les autres blocs tels que « parcelle », « sol » et « atmosphère ».

Les associations simples et les agrégations sont représentées de la même façon que les compositions, sauf que le rectangle du bloc apparaît en pointillé. Pour représenter plusieurs niveaux de compositions ou de références dans un même diagramme de bloc interne, il faut incorporer le bloc de niveau inférieur à l'intérieur du bloc englobant, tel que « végétaux » et « bétail » dans le bloc « parcelle ».

Un connecteur est un élément structurel utilisé pour relier deux parties et indiquer leur interaction. Les connecteurs permettent de relier finement les parties à travers des ports. Les multiplicités apparaissent aux extrémités d'un connecteur. Elles décrivent le nombre d'instances qui peuvent être mises en relation par le biais de liens décrits par le connecteur.

De manière générale, une association (e.g. une agrégation ou une composition) permet de relier des blocs, un connecteur ou des parties tandis qu'un lien permet de relier des instances.

Le diagramme de bloc interne permet également de décrire la logique de connexion, de services et de flots entre les blocs, grâce au concept de port. Les ports peuvent être de deux natures :

- Flux ou « flow port », qui autorise la circulation de flux physiques (matière, énergie ou information) entre les blocs. La nature du flux peut être de n'importe quel type : flots de données échantillonnées, signal physique, mais aussi énergie, ou encore fluide ; par exemple des effluents, de l'eau, etc.
- Port standard, qui permet la description de services logiques entre les blocs, au moyen d'interfaces regroupant des opérations (par exemple des commandes venant d'un agent).

Les « flow ports » sont soit composites (agrégation de flux de natures différentes comme par exemple l'épandage en surface), soit atomiques (un seul flux ; par exemple de l'eau pour l'irrigation). La direction du flux est indiquée par une flèche à l'intérieur du carré représentant le port. Les ports standards sont représentés par de simples carrés. Il est possible de typer les



« flow ports » par un bloc, un « value type » ou un signal, représentant le type d'élément pouvant circuler en entrée ou en sortie du port. Plusieurs ports peuvent avoir le même type.

Les éléments de flot (« item flows ») permettent de décrire ce qui circule réellement sur les connecteurs alors que les « flow ports » définissent simplement la nature de ce qui peut circuler. Quand un point d'interaction a une interface complexe avec plusieurs flux, le « flow port » correspondant doit être modélisé comme un « flow port » composite (ou non atomique). Dans ce cas, le port doit être typé par une spécification de flux (« flow specification »), définie dans un diagramme de définition de bloc. Elle inclut plusieurs propriétés de flux, chacune ayant un nom, un type et une direction (« in », « out » ou « inout ») - voir les définitions des « flow specification » pour l'épandage en surface et l'irrigation.

Un « flow port » composite est représenté par deux crochets se faisant face, dessinés à l'intérieur du symbole du port. Quand deux parties interagissent, elles échangent souvent les mêmes flux mais avec des directions opposées. Plutôt que de créer une deuxième « flow specification » en inversant toutes les directions des propriétés, SysML permet de réutiliser la même spécification en indiquant simplement que le deuxième « flow port » est conjugué. Les ports conjugués sont représentés graphiquement en inversant les couleurs (crochets blancs sur fond noir) - voir la définition de la « flow specification » pour l'épandage en surface.

Les interfaces sont des regroupements cohérents d'opérations placés sur des ports. Graphiquement, une interface est représentée comme un bloc avec le mot-clé «interface» ou le symbole d'un cercle dans le coin supérieur droit. Les deux types d'indication d'interfaces placées sur un port sont les interfaces fournies (« provided interface »), qui spécifient les opérations que le bloc fournit, et les interfaces requises (« required interface »), qui spécifient les opérations dont le bloc a besoin pour réaliser son comportement. Il est possible de préciser à quel port standard est rattachée chaque interface. Un même port peut posséder plusieurs interfaces requises ou fournies (par exemple pour distinguer plusieurs classes de fonctionnalités). Plusieurs ports peuvent être reliés aux mêmes interfaces. Voir l'exemple avec des interfaces de données et de commande entre une personne (l'exploitant agricole) et les machines d'irrigation des parcelles.

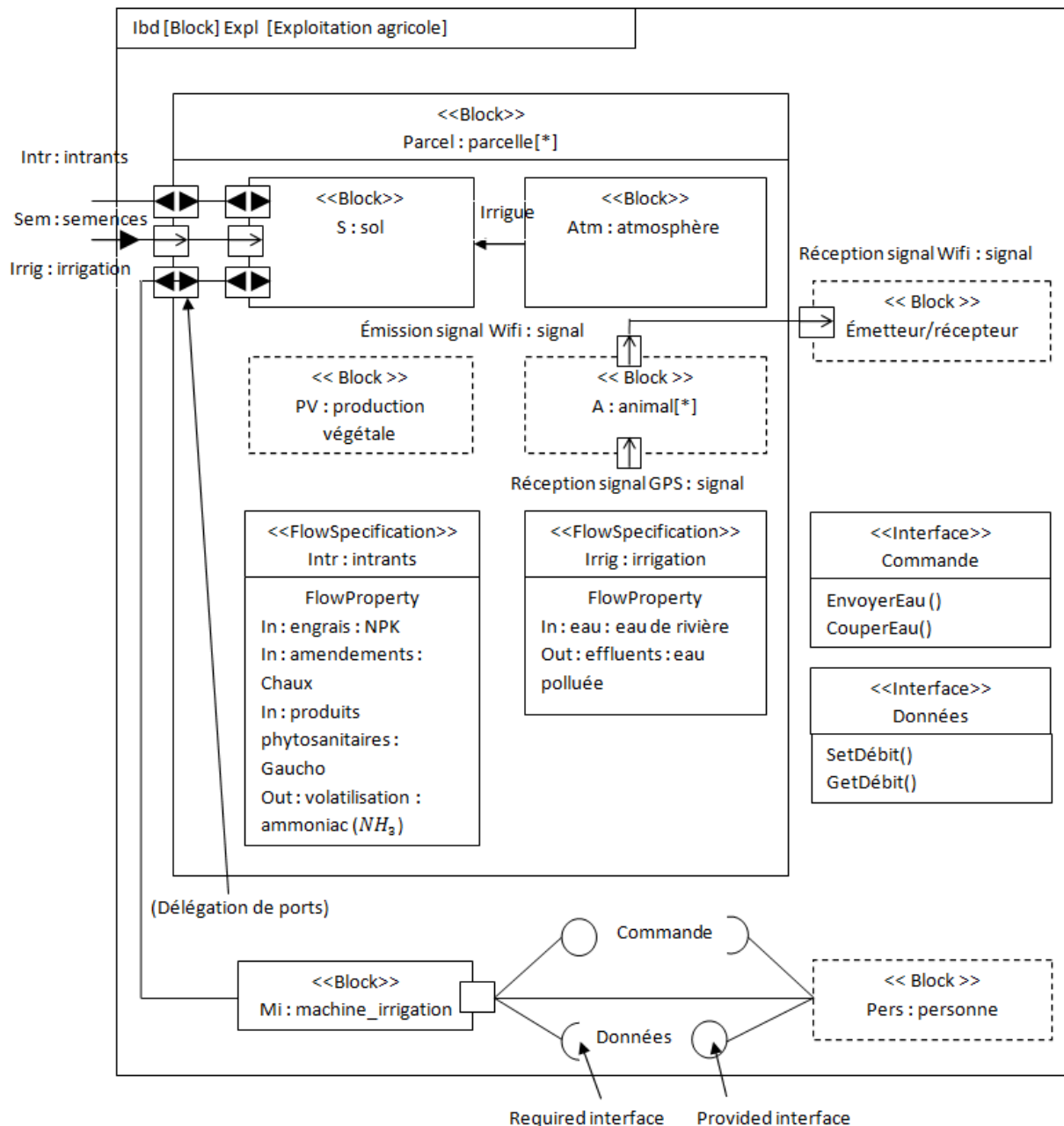


Figure 7 : exemple de diagramme de bloc interne modélisant un agrosystème

### 2.3.4 Modélisation des relations numériques et des contraintes sur les valeurs des paramètres d'un système

Le diagramme paramétrique permet de représenter des contraintes sur les valeurs des paramètres d'un système, ainsi que des relations numériques relatives à des aspects tels que les performances, la fiabilité, la masse, etc. Il s'agit d'un diagramme spécifique à SysML (spécialisation du diagramme de bloc interne). Les blocs modélisent alors des contraintes entre paramètres (et non plus des éléments physiques), ce qui permet de représenter

graphiquement des relations mathématiques ou des équations. Ce diagramme fournit également un support pour les études d'analyse système. Le diagramme paramétrique est composé de blocs de contraintes (dont le mot-clé est « constraint »), contenant des équations. L'expression de ces équations peut être soit informelle (i.e. du simple texte), soit formelle. Cependant, SysML ne fournit pas de langage d'expression mathématique. On exprime les relations dans un langage tels qu'OCL<sup>14</sup> ou encore MathML<sup>15</sup>.

Par exemple, dans le cadre de systèmes socio-environnementaux tels que des exploitations agricoles, des modèles de production permettent d'établir des relations entre la quantité d'engrais épanchée et la production végétale, des modèles économiques permettent de relier le cours des matières agricoles et la production avec les revenus des agriculteurs, etc.

Le bloc « équation de la variation de la masse de MOS » représente la différence entre les gains en matière organique du sol et les pertes. Le bilan s'écrit  $\frac{dC}{dt} = K_1.m - K_2.C$ . Le bloc « équation de la masse de MOS » exprime le fait que la variation de masse de MOS est la dérivée de la masse de MOS par rapport au temps. Le bloc « équation de la masse de MOS à l'équilibre » représente la quantité de MOS à l'équilibre ( $C_e = \frac{K_1.m}{K_2}$ ) et la quantité de MOS au temps  $t=0$  est représentée par la variable  $C_0$ . L'équation de la masse de MOS est :  $C = C_e + (C_0 - C_e). \exp(-K_2.t)$ .

Les relations entre blocs nécessitent des contraintes d'identité d'une variable d'un bloc à l'autre. Par exemple, le paramètre  $m$  (gain de MOS par hectare) dont il est question dans le bloc « équation de la masse de MOS à l'équilibre » est la même grandeur physique que le paramètre de même nom apparaissant dans le bloc « équation de la variation de la masse de MOS ». Le diagramme paramétrique peut également utiliser des « value binding » qui permettent partiellement une contrainte en fixant des valeurs numériques aux paramètres, comme par exemple la température moyenne annuelle de l'air (atm.temp en °C).

---

<sup>14</sup> OCL signifie Object Constraint Language. Il s'agit d'un langage associé à UML

<sup>15</sup> MathML signifie Mathematical Markup Language. Il s'agit d'un langage basé sur XML permettant l'affichage de symboles mathématiques (recommandation du W3C : <http://www.w3.org/TR/MathML3/>, page consultée le 01/03/2016.)

Le diagramme paramétrique ci-dessous illustre le modèle d'équation de Hénin-Dupuis [Hénin et Dupuis, 1945] avec des « value bindings » :

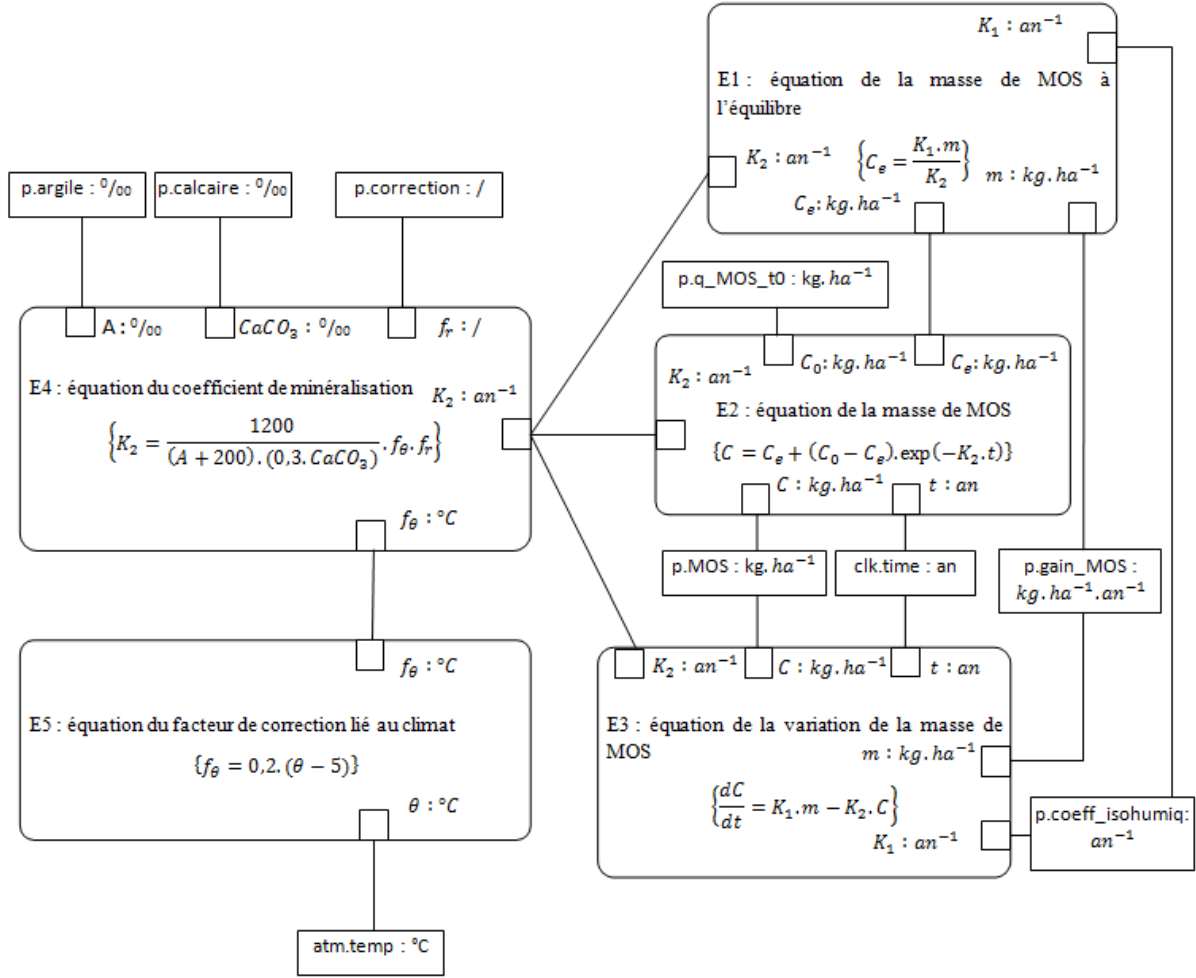


Figure 8 : diagramme paramétrique représentant les différentes équations du modèle de Hénin-Dupuis (légende des unités : kg = kilogramme ; a = année ; ha : hectare)

## 2.4 Simulation de systèmes

L'approche classique utilisée pour estimer les paramètres d'un système dynamique, lorsque les données d'entrée sont incertaines, est la méthode de Monte-Carlo (ou MMC). Cette approche consiste à réaliser un grand nombre de simulations avec tirages aléatoires des signaux d'entrée, en observant les sorties obtenues en fonction de ces entrées afin d'estimer les paramètres du système simulé. Il s'agit en fait d'une méthode d'analyse de sensibilité par tirages aléatoires qui se base sur les expériences répétées (simulations) pour l'estimation de valeurs et de paramètres de systèmes « complexes », en introduisant une approche statistique du risque. Cette méthode permet d'avancer toutes les possibilités liées à une observation, en

quantifiant l'incertitude statistique associée. Concrètement, ces techniques sont couramment utilisées dans divers domaines (physique, ingénierie, finance...), pour estimer la forme d'un signal, calculer la sensibilité d'un capteur, calculer les risques liés à une décision financière, estimer le prix d'une action, etc. Les outils logiciels Matlab et Simulink sont généralement utilisés pour réaliser de telles simulations. La méthode de Monte-Carlo est souvent la seule approche utilisable pour l'étude numérique des systèmes non linéaires de grande dimension pour lesquels aucune approche analytique n'est applicable.

## 2.5 Conclusion : les limites des approches actuelles de modélisation et simulation des systèmes

L'étude de la modélisation et la simulation de systèmes nous a permis d'identifier un certain nombre de limites concernant la modélisation standard de systèmes par SysML. En effet, SysML n'intègre pas de manière satisfaisante les notions d'incertitude sur les données, plus particulièrement floues et probabilistes. De fait, SysML prend simplement en compte des propriétés de blocs qui peuvent être déclarées comme des distributions de probabilité. Les données peuvent en effet être entachées d'incertitudes qui ne relèvent pas de distributions de probabilité (basées sur des répétitions d'événements) mais plutôt d'autres types d'incertitudes comme l'imprécision ou l'incomplétude, ce que ne permet pas la modélisation SysML. Par exemple, la notion de surface décrite dans l'exemple précédent peut être imprécise ou incomplète si le cadastre est mal renseigné ou inexistant.

Enfin, il n'y a pas de sémantique claire dans SysML, surtout en ce qui concerne la syntaxe graphique. En effet, un bloc peut potentiellement correspondre à des éléments de natures totalement différentes. En particulier, il n'y a pas de lien clair entre les diagrammes paramétriques (constitués de blocs qui implémentent des fonctions décrivant des contraintes entre paramètres) et les diagrammes de bloc (organes physiques), notamment en ce qui concerne la nature des ports (« flow ports ») et des valeurs (« value bindings »). Il serait également intéressant de pré-renseigner ces derniers avec des notions d'agrégats, afin d'automatiser les calculs après modélisation.

Enfin, les approches analytique et systémique doivent faire face aux défis du « big data ». Cette expression en vogue désigne des ensembles de données à la fois très volumineux et dynamiques qui nécessitent de travailler avec d'autres solutions que les outils de gestion de base de données classiques. Un défi majeur est donc de traiter à haute fréquence voire en

temps réel d'immenses quantités de données, de structures et de formats de données régis par des modèles complexes, de manière à obtenir des indicateurs d'aide à la décision en temps réel. De plus, il n'est pas toujours possible de nettoyer et d'optimiser les échantillons de données à temps sachant que les types d'imperfections entachant les données brutes sont variés : certains ne peuvent être basées ni sur les statistiques (étude de phénomènes qui se répètent dans le temps), ni sur les probabilités subjectives. En effet, une distribution de probabilité ne modélise pas de manière satisfaisante les degrés de vérité (exemple : « il fait presque beau temps ») ainsi que certains états de la connaissance d'un agent, qui peuvent être des informations à la fois incertaines, incomplètes, imprécises ou même linguistiques [Dubois et Prade, 1988]. Nous étudierons les approches de modélisation d'informations imparfaites dans la suite de cet état de l'art, nécessaires à la modélisation des systèmes socio-environnementaux.

L'approche qui consiste à simuler une variable aléatoire par des calculs numériques (méthode de Monte Carlo) possède elle aussi un certain nombre de limites. En effet, malgré le fait que la vitesse de convergence est indépendante de la dimension, simuler sous des lois de probabilité sur des espaces de grande dimension est un problème très difficile. De plus, cette approche nécessite un certain nombre de données ne permettant pas de prendre en compte l'ensemble des types d'incertitudes dont peuvent être entachées les données, telles que l'incomplétude ou encore l'imprécision sur les valeurs. Nous verrons dans la suite de cet état de l'art que certaines approches comme les approches possibilistes et l'arithmétique d'intervalle ont un coût computationnel bien moindre, tout en donnant pour résultats ce qui est réellement possible [Guillard et al., 2011] et en modélisant ou combinant de manière rigoureuse les caractères incomplet ou imprécis des données [Dubois et Prade, 1988] [Zadeh, 1978]. Nous verrons que ces méthodes concordent davantage avec notre objectif qui est d'envisager les scénarios croisés selon différentes hypothèses incertaines par un seul calcul formel, au lieu d'un grand nombre de calculs numériques.

### 3 Chapitre 3 : aide à la décision, modélisation et manipulation d'informations imparfaites

Dans le contexte des systèmes socio-environnementaux, le but des experts scientifiques est généralement de fournir des méthodes pour que le décideur prenne les meilleures décisions possibles afin d'optimiser la productivité et la durabilité de son entreprise. L'approche scientifique pour l'aide à la décision est appelée théorie de la décision. Il s'agit d'un domaine pluridisciplinaire qui est au carrefour de multiples disciplines telles que l'économie, la psychologie, la philosophie, les sciences de gestion, les mathématiques et l'informatique. De nos jours, de tels problèmes scientifiques nécessitent de tenir compte de caractéristiques d'incertitude, notamment l'incomplétude et l'imprécision des informations. C'est particulièrement le cas dans les sciences « molles » (e.g. la biologie, l'agronomie) où les variables observées ainsi que leurs relations peuvent être entachées d'incertitudes ou d'imprécision. L'approche analytique traditionnelle a développé des outils et des méthodes en supposant que l'incertitude est essentiellement aléatoire. Cette incertitude est généralement traitée au moyen de la théorie des probabilités. Cependant, d'autres approches telles que la théorie des sous-ensembles flous, la théorie des possibilités, la théorie de l'évidence et les *p*-boîtes généralisées se sont largement développées ces dernières années, élargissant la portée des périmètres des études scientifiques qui supposaient que l'incertitude est essentiellement aléatoire. Ces théories permettent ainsi de rendre compte des diverses formes d'imperfection des données (discrètes ou continues) rencontrées dans les systèmes socio-environnementaux.

Dans ce chapitre, nous présenterons les principales méthodes d'aide à la décision généralement utilisées dans le cadre des systèmes socio-environnementaux et qui nous seront nécessaires dans la suite de nos travaux. Nous présenterons ensuite quelques applications dans le domaine de l'agriculture, plus particulièrement des approches d'évaluation multicritères de systèmes agricoles. Enfin, nous présenterons les approches mathématiques de modélisation et de manipulation d'informations imparfaites.

## 3.1 Aide à la décision

### 3.1.1 Choix et préférences

La théorie de la décision a pour objet la prise de décision par un agent (décideur). Le choix de l'agent fait intervenir des arbitrages entre ses préférences (ou objectifs), afin de prendre la « meilleure » décision possible. Un choix correspond en général à une conjonction de caractéristiques d'un objet, de prévisions, etc. pouvant chacune prendre plusieurs valeurs. L'ensemble des valeurs que peuvent prendre ces caractéristiques est appelé domaine. Soit  $X$  un ensemble de choix possibles. Soit  $(x, y) \in X^2$ .

**Définition (relation de préférence faible) :** on appelle relation de préférence faible une relation binaire  $\geq$  sur  $X^2$ .  $x \geq y$  signifie que  $x$  est préféré ou indifférent à  $y$  [Abdellaoui et Gonzales, 2009].

Remarque : le fait que deux éléments  $x$  et  $y$  soient incomparables est traduit par :  $(Non(x \geq y) \text{ et } Non(y \geq x))$

**Définition (relation de préférence forte ou stricte) :** le fait de préférer strictement  $x$  à  $y$  se traduit par  $x > y$ .

**Définition (relation d'indifférence) :** lorsque l'agent est indifférent entre  $x$  et  $y$ , on écrit  $x \sim y$ .

Pour pouvoir manipuler les relations définies ci-dessus, il faut pouvoir quantifier et comparer les choix possibles à l'aide d'une relation entre les éléments d'un ensemble. On pourra ainsi exprimer les préférences. L'approche classique est d'ordonner les choix possibles selon le gain qu'ils apportent. On peut avoir :

- Un ordre total, i.e. une relation d'ordre binaire réflexive, transitive et antisymétrique où tous les choix sont comparables. Il n'y a pas de relation d'équivalence possible entre deux choix.
- Un pré-ordre total, i.e. relation binaire réflexive et transitive où tous les choix sont comparables. La relation d'équivalence est alors notée  $\sim$ .
- Un ordre partiel, i.e. tous les éléments ne sont pas comparables.

Prenons l'exemple d'un agriculteur qui a trois choix possibles de cultures dans une parcelle (décision monocritère) : blé, maïs ou colza. Si l'agriculteur pense «  $blé > colza > maïs$  »,



on a alors un pré-ordre total. S'il pense « *blé*  $\geq$  *colza*  $\geq$  *maïs* », on a alors un pré-ordre total. S'il pense « *blé*  $>$  *maïs* », « *blé*  $>$  *colza* » et que blé n'est pas comparable à colza, on a un ordre partiel.

Toutefois, sans représentation numérique des « gains » à l'aide de fonctions d'utilité, les relations  $\geq$ ,  $>$  et  $\sim$  sont très difficiles à manipuler.

### 3.1.2 Les fonctions d'utilité

Les travaux de Raiffa [Raiffa, 1969] et Edwards [Edwards, 1971] sont parmi les premiers concernant l'aide à la décision. Le modèle formel proposé fait appel à des représentations numériques (appelées scores ou utilités) des préférences de l'agent relatives aux actions possibles<sup>16</sup> par le biais de fonctions numériques appelées fonctions d'utilité. Les utilités permettent de classer les choix possibles du plus désirable (« meilleur » choix) au moins désirable.

La construction d'une fonction d'utilité nécessite (1) de pouvoir représenter numériquement les préférences du décideur (conditions de cohérence) et (2) de pouvoir agréger de manière simple des fonctions d'utilité mono-attributs en fonctions d'utilité multi-attributs qui sont alors décomposables. Ces dernières sont les plus courantes car les attributs – ou objectifs – sont en général multiples. Chaque agent ayant ses propres préférences et objectifs, il existe potentiellement autant de fonctions d'utilité que de décideurs. La construction d'une fonction d'utilité peut devenir très complexe étant donné que ces objectifs peuvent parfois être contradictoires. Par exemple, il peut être difficile de concilier des critères comme « gagner de l'argent » ou bien « être productif » avec « créer du bien-être » ou « préserver la nature ». Une fonction d'utilité se modélise formellement de la manière suivante :

Soit une fonction  $u: X \rightarrow \mathbb{R}$ . On a :  $\forall (x, y) \in X^2, x \geq y \Leftrightarrow u(x) \geq u(y)$ .

La fonction  $u$  attache à tout choix possible de l'ensemble  $X$  un nombre réel, de telle sorte que plus ce dernier est élevé, plus le choix est censé être « préféré » par l'agent. En pratique, le décideur a des objectifs multiples, ce qui conduit à décrire les conséquences possibles grâce à de multiples attributs, i.e. l'ensemble des conséquences d'un choix est multidimensionnel.

---

<sup>16</sup> Dans la suite de ce document, nous désignerons indifféremment les actions possibles par « hypothèses », « choix » et alternatives ».

Dans le cas d'objectifs multiples, une fonction d'utilité se modélise formellement de la manière suivante :

$$\forall x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in X^2, x \geq y \Leftrightarrow u(x_1, \dots, x_n) \geq u(y_1, \dots, y_n)$$

La construction d'une fonction d'utilité multi-attributs est difficile, c'est pourquoi on cherche en général à décomposer celle-ci en fonction d'utilité mono-attributs, plus faciles à construire individuellement.

La forme la plus fréquente est la décomposition polynômiale multilinéaire de fonctions d'utilités mono-attributs. Dans ce cas, une fonction d'utilité se modélise formellement de la manière suivante :

$$\exists u_i: X_i \rightarrow \mathbb{R}, \forall j \in J (\text{avec } J = 2^{\{1, \dots, n\}}), \exists \pi_j \in \mathbb{R} / u(x_1, \dots, x_k) = \sum_{j \in J} \pi_j \prod_{k \in j} u_k(x_k)$$

Remarque :  $J = 2^{\{1, \dots, n\}}$  est l'ensemble des parties de  $\{1, \dots, n\}$ .

Remarque : des cas particuliers de cette forme sont la décomposition additive ( $u(x_1, \dots, x_k) = \sum_{k=1}^n u_k(x_k)$ ), la décomposition multiplicative ( $u(x_1, \dots, x_k) = \prod_{k=1}^n u_k(x_k)$ ) et la décomposition linéaire  $u(x_1, \dots, x_k) = \sum_{k=1}^n \pi_j \cdot u_k(x_k)$ , qui correspondent respectivement à la somme, au produit et à la combinaison linéaire de fonction d'utilités.

### 3.1.3 Décisions en contexte incertain

Trois concepts de base sont présents dans la théorie de la décision :

- Les alternatives. En théorie de la décision, c'est en général un ensemble fini de choix mutuellement exclusifs.
- Les événements (ou états de la nature). Ce sont des facteurs en général incertains qui peuvent modifier les conséquences d'une décision.
- Les conséquences (résultats d'une décision). Les valeurs des résultats correspondent en général à des utilités.

Une décision dans le certain signifie qu'une décision donnée a toujours la même conséquence. Dans ce cas, dans le cadre des systèmes socio-environnementaux, une même dose d'engrais épandue sur un champ donnerait toujours la même production végétale, ce qui est rarement le cas. En effet, des événements peuvent modifier les conséquences d'une décision. Dans ce cas, chaque alternative peut avoir plusieurs conséquences possibles, en fonction de la réalisation

de tel ou tel événement. En théorie de la décision individuelle, la représentation standard d'un problème de décision est la matrice de décision. Une autre représentation d'un problème de décision est l'arbre de décision.

Exemple : un agriculteur a deux possibilités : cultiver du maïs et cultiver du blé sur une parcelle donnée. On envisage plusieurs événements météo avec une fonction d'utilité (qui est par exemple proportionnelle à la rentabilité) : (1) si la météo est bonne, alors cultiver du blé a une utilité de 100, et cultiver du maïs a une utilité de 150 ; (2) si la météo est mauvaise, cultiver du blé a une utilité de 80, tandis que cultiver du maïs a une utilité de 50 seulement (le maïs est plus sensible aux aléas météorologiques). La matrice de décision s'écrit :

Tableau 3 : matrice de décision

Décision \ événement	Bonne météo	Mauvaise météo
Blé	100	80
Maïs	150	50

Dans la suite de ce chapitre, nous nous baserons sur cet exemple très simple pour illustrer les différentes approches d'analyse de la décision.

### 3.1.4 Analyse de la décision dans le risque

L'analyse de la décision consiste à définir des méthodologies pour l'aide à la décision. Dans certains cas, les distributions de probabilité des événements sont connues (probabilités objectives). Ce type d'analyse est aussi appelé analyse de la décision à base de loterie. En effet, dans le cas d'une loterie, les probabilités de gagner ainsi que les gains correspondants sont connus à l'avance. Les théorèmes sur les probabilités permettent de déterminer les probabilités d'événements plus complexes (analyse combinatoire). Par exemple, la probabilité d'un gain lors de  $n$  tirages indépendants à la loterie est le produit des  $n$  probabilités des gains à chaque tirage de loterie. Enfin, les méthodes statistiques (échantillonnage, sondages, tests) sont des moyens puissants pour estimer des probabilités d'événements.

A chaque alternative est assignée une moyenne pondérée de ses valeurs d'utilité sous différents états de nature, et les probabilités de ces états sont utilisées comme poids. Supposons qu'une décision corresponde à la loterie  $(x_1, p_1; \dots; x_n, p_n)$ , i.e. cette décision va avoir pour conséquence  $x_1$  avec une probabilité  $p_1$ , ...,  $x_n$  avec une probabilité  $p_n$ . Soit  $x_j$  une décision possible et soit  $u(x_{i,j})$  l'utilité du scénario « décision  $x_j$  et événement  $i$  ».

L'axiomatique de Von Neumann-Morgenstern [Von Neumann et Morgenstern, 1944] implique alors l'existence d'une utilité espérée (UE) d'une décision  $x_j$  dont l'expression est :

$$UE(x_j) = \sum_{i=1}^n p_i \cdot u(x_{i,j})$$

Dans notre exemple, supposons que les probabilités d'une bonne et d'une mauvaise météo sont 0,5. Après calcul, les UE de chaque décision sont :

- $UE(blé) = 90$
- $UE(maïs) = 100$

L'agriculteur déciderait de cultiver du maïs qui a une utilité espérée (UE) de 100.

Dans le cas où l'on ne suppose pas l'existence d'une distribution de probabilité objective sur les événements, les fonctions d'utilité impliquent l'existence de croyances de l'agent sur les différents états du monde. Ces croyances sont classiquement modélisées par des probabilités subjectives. On parle alors d'incertain mesurable. Par exemple, dans le domaine de l'agriculture, le décideur fait souvent des choix (par exemple cultiver tel ou tel type de culture : maïs, blé, etc.) sans connaître de manière certaine la probabilité de certains événements comme les aléas météo, les catastrophes sanitaires, les modifications de l'état du sol, etc. En effet, l'approche objectiviste des probabilités ne peut s'appliquer car l'agriculteur n'a pas à disposition tous les résultats des scénarios possibles de l'exploitation en termes productifs, agronomiques, économiques, écologiques, etc. Dans ce cas, le décideur associe à chaque état de la nature une croyance  $\pi_i$ . L'utilité espérée de  $x_j$  est, comme dans le modèle de von Neumann-Morgenstern :

$$UE(x_j) = \sum_{i=1}^n \pi_i \cdot u(x_{i,j})$$

En incertain mesurable, on utilise donc des probabilités subjectives modélisant les croyances de l'agent décideur pour pondérer les importances respectives de chaque état. Dans ce cas, pour choisir la décision la plus avantageuse, l'agent décideur peut se baser sur les informations de probabilités disponibles. Le critère principal pour la décision en incertain mesurable est le critère de Pascal, qui consiste à prendre la décision qui correspond au maximum de l'utilité espérée (somme des utilités pondérées par les probabilités). Dans certains cas, le critère de Pascal peut ne pas convenir du fait de l'aversion au risque du

décideur. En effet, l'utilité des gains peut ne pas être appréciée de la même manière par des agents différents. Cette théorie est très largement utilisée en économie [Marshall, 1890]. Dans le cas où la fonction d'utilité de la richesse a une forme concave (dérivée positive et dérivée seconde négative ; par exemple la fonction d'utilité logarithmique de Bernoulli), l'utilité des gains croît d'autant moins vite que l'agent est riche au départ, ce qui fait qu'un agent « pauvre » aura tendance à prendre moins de risques pour sécuriser ses gains (aversion au risque). En revanche, une fonction d'utilité de la richesse convexe (dérivée positive et dérivée seconde positive) traduit un goût pour le risque.

### 3.1.5 Analyse de la décision en contexte incertain

Dans ce paragraphe, on envisage un scénario de décision en contexte incertain non mesurable sous forme d'arbre avec plusieurs hypothèses possibles. Dans ce cas, on suppose qu'il n'existe pas de distribution de probabilité sur les événements ayant un effet sur les conséquences et donc sur l'utilité des scénarios. Comme nous l'avons déjà précisé dans les paragraphes précédents, le but recherché par l'agent est de prendre la meilleure décision. Pour cela, il doit envisager les scénarios possibles, et choisir une hypothèse en fonction de critères.

Soit  $x_j$  une décision possible et soit  $u(x_{i,j})$  l'utilité du scénario « décision  $x_j$  et événement  $i$  ».

Dans la suite de ce paragraphe, nous présentons les principaux critères existants pour la décision en incertain non mesurable [Espinasse, 2010] :

- Le critère du maximax consiste à choisir la décision qui a la plus grande utilité maximale. C'est en fait le critère du décideur optimiste. Dans notre exemple, l'agriculteur déciderait de cultiver du maïs qui a une utilité de 200 dans le meilleur des cas.
- Le critère du minimax (également appelé critère de Wald) consiste à choisir la décision qui a la plus grande utilité minimale. C'est en fait le critère du décideur pessimiste. Dans notre exemple, l'agriculteur déciderait de cultiver du blé.
- Le critère du regret minimax (également appelé critère de Savage) Savage consiste à calculer une « matrice des regrets ». Le regret correspond au « manque à gagner » d'une décision par rapport au meilleur scénario. Il est calculé en soustrayant le maximax et l'utilité de chaque scénario possible. Le décideur prend alors la décision qui correspond au plus petit regret maximal. Dans notre exemple, les regrets sont les suivants : blé : 50 (si la météo est bonne) et 70 (si la météo mauvaise) ; maïs : 0 (si la météo est bonne) et 100 (si la météo est

mauvaise). Dans ce cas, l'agriculteur décidera donc de cultiver du blé (qui a le regret maximal le plus petit).

- Le critère de Laplace consiste à calculer l'utilité moyenne conséquence à chaque décision possible. L'agent prend la décision qui correspond à l'utilité moyenne la plus grande. Ce critère est fondé sur l'hypothèse d'équiprobabilité des événements possibles (soit  $1/n$  pour chaque état s'il y a  $n$  états). Le critère de Laplace s'écrit :

$$L(x_j) = \frac{1}{n} \sum_{i=1}^n u(x_{i,j})$$

Dans notre exemple, l'agriculteur déciderait de cultiver du maïs qui a une « utilité moyenne » de 100.

- Le critère d'Hurwicz consiste à prendre la pire et la meilleure utilité de chaque scénario et de les pondérer respectivement par un degré de pessimisme  $\alpha$  et un degré d'optimisme  $(1 - \alpha)$ ,  $\alpha$  étant choisi arbitrairement dans  $[0,1]$ . Le décideur choisit la décision qui donne la plus grande valeur de ce critère. Le critère d'Hurwicz s'écrit :

$$H(x_j) = \alpha \cdot \min(u(x_{i,j})) + (1 - \alpha) \cdot \max(u(x_{i,j}))$$

Dans notre exemple, avec  $\alpha$ , l'agriculteur déciderait de cultiver du maïs qui a une « utilité agrégée » de 100.

- Le critère de Bernouilli consiste à calculer la moyenne du logarithme conséquence à chaque décision possible. L'agent prend la décision qui correspond au critère le plus grand. Comme le critère de Laplace, ce critère est basé sur l'hypothèse d'équiprobabilité. Le critère de Laplace s'écrit :

$$B(x_j) = \frac{1}{n} \sum_{i=1}^n \log(u(x_{i,j}))$$

On obtient  $B(\text{blé}) = 3.903$  et  $B(\text{maïs}) = 3.875$ . L'agriculteur déciderait donc de cultiver du blé.

Pour une décision optimale, on peut construire un tableau classant les décisions de l'utilité par critère et faisant la somme des rangs [Espinasse, 2010] :

Tableau 4 : tableau des rangs par critère de décision

Décision \ critère	Maximax	Minimax	Regret minimax	Laplace	Hurwicz	Bernoulli	Somme des rangs
Blé	2	1	1	2	2	1	9
Maïs	1	2	2	1	1	2	9

Cet exemple illustre bien le fait qu'il est difficile de trouver un critère approprié. En effet, on ne peut pas classer les deux décisions car leur somme des rangs est égale.

### 3.1.6 Agrégation de décisions multicritères

Dans le cadre d'une décision, les choix possibles peuvent être décrits par plusieurs critères, avec des relations de préférences différentes. Il arrive que les critères ne soient pas « agréables » en une unique valeur ou qu'ils soient conflictuels, ce qui rend délicat la comparaison des différents choix.

Soit  $X$  un ensemble d'alternatives possibles. Chaque élément de  $X$  est un vecteur d'alternatives. On a donc  $X_1 \times \dots \times X_n$  ( $n$  critères ayant chacun leur fonction d'utilité et leur relation de préférence).

Nous avons identifié deux types principaux d'approches pour la décision multicritère :

- Les approches d'agrégation quantitatives. Dans cette approche, on agrège les critères en un critère « de synthèse » pour chaque choix possible. Enfin, on compare les critères « de synthèse » obtenus [Marichal, 2009].
- Les approches qualitatives, e.g. les méthodes ELECTRE [Roy, 1968] [Roy, 1996] et PROMETHEE [Brans et Mareschal, 2002]. Dans cette approche, on compare les choix critère par critère puis on agrège les résultats des comparaisons.

Dans la suite de ce chapitre, nous traiterons uniquement les approches d'agrégation quantitatives.

- Soient  $x_1 = (x_1^1, \dots, x_1^n)$  et  $x_2 = (x_2^1, \dots, x_2^n) \in X$  avec  $X = X_1 \times \dots \times X_n$ .  $x_1$  et  $x_2$  sont deux choix évalués sur  $n$  critères).
- Soit  $\geq_i$  une relation de préférence sur le critère  $i$ .

- Soit  $\tau$  une permutation telle que  $x_1^{\tau(1)} \leq \dots \leq x_1^{\tau(n)}$
- Soit  $v$  une permutation telle que  $x_2^{v(1)} \leq \dots \leq x_2^{v(n)}$

### 3.1.6.1 Dominance de Pareto

Définition (dominance de Pareto) : on dit que  $x_1$  domine  $x_2$  au sens de Pareto ssi :

$$\forall i \in [1, n], x_1^{\tau(i)} \geq x_2^{v(i)} \text{ et } \exists i \in [1, n] / x_1^{\tau(i)} > x_2^{v(i)}$$

On note alors :  $x_1 >_p x_2$ .

Définition (Pareto optimalité) : on dit qu'un vecteur d'alternatives  $x_1 \in X$  est optimal au sens de Pareto ssi il n'existe pas de  $x_2 \in X$  tel que  $x_2 >_p x_1$

### 3.1.6.2 Dominance de Lorentz

Définition (dominance de Lorentz) : soient  $(u_1, u_2, \dots, u_n)$  les fonctions d'utilité associées aux critères  $X_1 \times \dots \times X_n$ . On dit que  $x_1$  domine  $x_2$  au sens de Lorentz ssi :

$$\forall i \in [1, n], \sum_{j=1}^i u_j(x_1^{\tau(j)}) \geq \sum_{j=1}^i u_j(x_2^{v(j)})$$

### 3.1.6.3 Moyenne des poids ordonnés (OWA<sup>17</sup>) [Yager, 1988]

Définition (préférence au sens OWA) : soient  $(u_1, u_2, \dots, u_n)$  les fonctions d'utilité associées aux critères  $X_1 \times \dots \times X_n$ . Soient  $(\omega_1, \omega_2, \dots, \omega_n)$  un jeu de poids. On dit que  $x_1$  est préféré à  $x_2$  ssi :

$$\forall i \in [1, n], \sum_{i=1}^n \omega_i \cdot u_i(x_1^i) \geq \sum_{i=1}^n \omega_i \cdot u_i(x_2^i) \text{ avec } u_i(x_1^i) \leq u_i(x_2^i)$$

Cas particuliers :

- Le jeu de poids  $(1, 0, \dots, 0)$  correspond à l'opérateur min.
- Le jeu de poids  $(0, \dots, 0, 1)$  correspond à l'opérateur max.
- Le jeu de poids  $(1, \dots, 1)$  correspond à l'opérateur moyenne.

---

<sup>17</sup> Ordered Weight Average



#### 3.1.6.4 Médiane des poids

Définition (préférence au sens de la médiane) : on dit que  $x_1$  est préféré à  $x_2$  ssi :

$$\text{méd}\left(u_i(x_1^i)\right) \geq \text{méd}\left(u_i(x_2^i)\right)$$

#### 3.1.6.5 Intégrale de Choquet [Choquet, 1953]

Définition (préférence au sens de l'intégrale de Choquet) : On dit que  $x_1$  est préféré à  $x_2$  ssi :

$$\forall i \in [1, n], \sum_{i=1}^n w_{\{i, \dots, n\}} \cdot \left(u_i(x_1^i) - u_i(x_1^{i-1})\right) \geq \sum_{i=1}^n w_i \cdot u_i(x_2^i) \text{ avec } \forall i, u_i(x_1^{i-1}) \leq u_i(x_1^i)$$

#### 3.1.6.6 Conclusion sur les méthodes d'agrégation

Les méthodes d'agrégation présentent l'avantage de produire un optimum (pré-ordre total). Les méthodes de calcul sont simples et elles fournissent une grande variété de méthodes. Toutefois, ces méthodes additives présentent l'inconvénient d'être compensatoires et qu'il est de ce fait difficile de donner la priorité à un critère particulier. Enfin, ces méthodes obligent à comparer des critères de natures différentes, ce qui rend parfois les interprétations délicates en fonction des nombreuses méthodes pouvant être utilisées [Rolland, 2013].

### 3.1.7 Décisions à partir d'informations imparfaitement définies

Comme nous l'avons vu dans l'état de l'art, les informations imparfaitement définies peuvent être représentées de manière simple via la théorie des possibilités et celle des sous-ensembles flous : l'incertitude épistémique, l'imprécision (incomplétude et logique floue, par exemple modélisation de variables linguistiques). Pour composer des propositions floues, on utilise la logique floue qui est une extension de la logique booléenne. Nous présenterons en premier lieu les opérations ensemblistes de base permettant de calculer les fonctions d'appartenance résultant d'une conjonction ou d'une disjonction de propositions floues. En second lieu, nous étudierons les opérateurs d'agrégation qui, étendus aux sous-ensembles flous, permettent de réaliser un « consensus » entre variables et ainsi de prendre des décisions.

#### 3.1.7.1 Opérations ensemblistes

Les sous-ensembles flous sont décrits par leur fonction d'appartenance. Il s'agit généralement d'une fonction trapézoïde continue de  $\mathbb{R}$  vers  $[0,1]$  définie par un noyau et un support. Pour combiner de telles variables et en déduire leur fonction d'appartenance, on réalise des

opérations ensemblistes sur les sous-ensembles flous telles que l'union et l'intersection. Les fonctions d'appartenance résultantes de telles opérations sont définies grâce à une t-norme (norme triangulaire) et une t-conorme.

Une t-norme est une fonction  $T: [0,1]^2 \rightarrow [0,1]$  qui vérifie les propriétés suivantes  $\forall (x, y, z, t) \in [0,1]^4$  :

- $T(x, y) = T(y, x)$  (commutativité)
- $T(x, T(y, z)) = T(T(x, y), z)$  (associativité)
- $x \leq z$  et  $y \leq t \Rightarrow T(x, y) \leq T(z, t)$  (monotonie)
- $T(x, 1) = x$  (élément neutre égal à 1)

Une t-norme  $T$  correspond à un opérateur d'intersection entre deux sous-ensembles flous. En effet, si  $A$  et  $B$  sont deux sous-ensembles flous sur  $\mathbb{R}$ , on en définit un troisième appelé  $C$ , qui est l'intersection de  $A$  et  $B$ . Soit  $x \in \mathbb{R}$ . On a :

$$f_C(x) = T(f_A(x), f_B(x))$$

De la même manière, on définit une t-conorme comme une fonction  $T': [0,1]^2 \rightarrow [0,1]$  qui vérifie les propriétés suivantes  $\forall (x, y, z, t) \in [0,1]^4$  :

- $T'(x, y) = T'(y, x)$  (commutativité)
- $T'(x, T'(y, z)) = T'(T'(x, y), z)$  (associativité)
- $x \leq z$  et  $y \leq t \Rightarrow T'(x, y) \leq T'(z, t)$  (monotonie)
- $T'(x, 0) = x$  (élément neutre égal à 0)

Une t-conorme  $T'$  correspond à un opérateur d'union entre deux sous-ensembles flous. En effet, si  $A$  et  $B$  sont deux sous-ensembles flous sur  $\mathbb{R}$ , on en définit un troisième appelé  $D$ , qui est l'union de  $A$  et  $B$ . Soit  $x \in \mathbb{R}$ . On a :

$$f_D(x) = T'(f_A(x), f_B(x))$$

Les t-normes et les t-conormes les plus courantes sont présentées dans le tableau suivant :

Tableau 5 : t-normes et t-conormes les plus utilisées (source : [Bouchon-Meunier, 1993])

Nom	t-norme (intersection)	t-conorme (union)
Zadeh	$\min(x, y)$	$\max(x, y)$
Probabiliste	$xy$	$x + y - xy$

Lukasiewicz	$\max(x, y - 1, 0)$	$\min(x + y, 1)$
Hamacher (avec $\gamma > 0$ )	$\frac{xy}{\gamma + (1 - \gamma)(x + y - xy)}$	$\frac{x + y - xy - (1 - \gamma)xy}{1 - (1 - \gamma)xy}$
Drastique	$x$ si $y = 1$ $y$ si $x = 1$ 0 sinon	$x$ si $y = 0$ $y$ si $x = 0$ 1 sinon

L'opérateur logique correspondant au complément (« non ») a pour valeur de vérité  $1-x$ .

Avec les opérations ensemblistes que nous venons d'étudier, il est possible de composer des données numériques imprécises et/ou des variables linguistiques. Une variable linguistique est représentée par un triplet  $(V, X, T_v)$  [Bouchon-Meunier, 1993] dans lequel  $V$  est une variable (e.g. la température) définie sur un ensemble  $X$  (e.g.  $\mathbb{R}$ ), et  $T_v$  est un ensemble fini ou infini de sous-ensembles flous de  $X$ . Soient deux variables linguistiques  $(V, X, T_v)$  et  $(V, X, T_w)$ . On peut ainsi combiner des variables linguistiques avec des opérateurs de conjonction (intersection « et ») et de disjonction (union « ou »).

### 3.1.7.2 Implications floues

Soit une règle floue de la forme « si  $V$  est  $A$  alors  $W$  est  $B$  ». Dans ce cas, le problème est de savoir comment inférer la règle si «  $V$  est à peu près  $A$  ». Soit  $(x, y) \in X \times Y$ . La variable floue  $V$  appartient à l'ensemble flou  $A$  avec un degré de validité  $f_A(x)$ . La variable floue  $W$  appartient à l'ensemble flou  $B$  avec un degré qui dépend donc du degré de validité  $f_A(x)$  de la prémisse. On définit alors la fonction d'appartenance  $f_R$  de la relation d'implication floue  $R$  entre  $X$  et  $Y$  qui s'exprime en fonction des fonctions d'appartenance  $f_A(x)$  de la prémisse et  $f_B(y)$  de la conclusion :

$$f_R(x, y) = \Phi(f_A(x), f_B(y))$$

La fonction  $\Phi$  n'est pas unique. Elle est définie de telle sorte que si  $A$  et  $B$  sont précis, l'implication floue est identique à l'implication de la logique booléenne classique. Le tableau suivant illustre les fonctions d'implications floues les plus courantes :

Tableau 6 : valeurs de vérité des différents types d'implication floue (source : [Bouchon-Meunier, 1993])

Nom	Valeur de vérité de l'implication floue
-----	---

Reichenbach	$1 - f_A(x) + f_A(x) \cdot f_B(y)$
Willmott	$\max(1 - f_A(x), \min(f_A(x), f_B(y)))$
Rescher-Gaines	$1 \text{ si } f_A(x) \leq f_B(y)$ $0 \text{ sinon}$
Kleene-Dienes	$\max(1 - f_A(x), f_B(y))$
Brouwer-Gödel	$1 \text{ si } f_A(x) \leq f_B(y)$ $f_B(y) \text{ sinon}$
Goguen	$\min(\frac{f_B(y)}{f_A(x)}, 1) \text{ si } f_A(x) \neq 0$ $1 \text{ sinon}$
Lukasiewicz	$\min(1 - f_A(x), f_B(y), 1)$
Mamdani	$\min(f_A(x), f_B(y))$
Larsen	$f_A(x) \cdot f_B(y)$

La prémisse d'une règle floue peut être formée de conjonctions et disjonctions de propositions floues. L'ensemble des règles ainsi obtenues est appelée matrice des décisions. Chaque règle est calculée grâce aux valeurs de vérité de son implication floue ci-dessus. Enfin, l'ensemble des résultats obtenus sont agrégés en un résultat final grâce à une t-conorme (opérateur « ou »). On obtient alors un sous-ensemble flou correspondant au résultat final. Il reste à prendre la décision finale en agrégeant ce sous-ensemble flou en une valeur unique.

### 3.1.7.3 Les opérateurs d'agrégation en logique floue

Un opérateur d'agrégation est une fonction qui permet de passer d'un ensemble de valeurs plus ou moins précises ou certaines à une décision unique (i.e. trouver une valeur exacte, un « consensus » entre valeurs floues) [Gacogne, 1997]. Ce processus est appelé « défuzzification ». Il existe un grand nombre d'opérateurs d'agrégation floue. Ils sont basés sur le principe suivant : soit  $x$  une variable définie sur un ensemble  $X$  (e.g.  $\mathbb{R}$ ). Soient  $A$  et  $B$  deux ensembles flous définis sur  $X$ . L'agrégation  $H(A, B)$  a pour forme :

$$f_{H(A,B)}(x) = H(f_A(x), f_B(x))$$

Les deux principales méthodes de défuzzification sont la méthode moyenne des maxima (MM) et la méthode du centre de gravité (COG).

La défuzzification COG (centre de gravité) [Mamdani, 1974] définit la sortie comme correspondant à l'abscisse du centre de gravité de la surface de la fonction d'appartenance caractérisant l'ensemble flou issu de l'agrégation des conclusions. En d'autres termes, c'est l'abscisse du centre de gravité de la surface sous la courbe résultat.

Soit  $x \in \mathbb{R}$ . Soit  $C$  l'ensemble flou issu de l'agrégation des conclusions.

$$Décision = \frac{\int_C x \cdot f(x) \cdot dx}{\int_C f(x) \cdot dx}$$

L'inconvénient de cette méthode est sa lenteur d'exécution. Pour palier ce problème, Sugeno [Sugeno, 1985] a proposé un opérateur similaire si ce n'est qu'il consiste à calculer la sortie comme la somme pondérée des conclusions de chaque règle (milieu du noyau de l'ensemble flou correspondant) dont le poids est le degré de vérité de la règle.

La défuzzification MM (moyenne des maximums) définit la sortie comme étant la moyenne des abscisses des maxima de l'ensemble flou issu de l'agrégation des conclusions. En d'autres termes, c'est la moyenne des valeurs de sortie les plus vraisemblables.

Soit  $S = \{x_m \in C / f(x_m) = \sup_{x \in C} (f(x))\}$

$$Décision = \frac{\int_S x \cdot f(x) \cdot dx}{\int_S dx}$$

#### 3.1.7.4 Décisions dans le cadre de la théorie de Dempster-Shafer

La probabilité pignistique est une mesure utilisée pour l'aide à la décision : la fonction de masse  $m$  est alors convertie en fonction de probabilité définie sur  $\Omega$ . Quelque soit le singleton  $\omega \in \Omega$ , l'expression de la probabilité pignistique est :

$$\text{BetP}(\omega) = \frac{1}{1 - m_\cap(\phi)} \sum_{\omega \in A} \frac{m(A)}{|A|}$$

avec  $|A|$  étant le cardinal du sous-ensemble  $A$ . La décision  $\omega$  est prise lorsque  $\text{BetP}(\omega)$  est maximum.

### 3.1.8 Applications dans le domaine de l'agriculture

Dans ce domaine, les approches OAD<sup>18</sup> sont des combinaisons de résultats des approches de modélisation précédemment présentées (approches analytique et systémique) ainsi que les fruits d'analyses à grande échelle basées sur des opinions d'experts (approche empirique). Ces outils d'aide à la décision sont des applications des approches mathématiques que nous venons d'étudier.

Il s'agit en général d'évaluations multicritères du système à partir de données statiques, en vue de prendre des décisions le concernant. Dans une telle approche, les indicateurs sont établis à partir des observations actuelles du système. Ils évaluent pour la plupart la durabilité du système et sont censés être représentatifs de la durabilité future du système. Les règles régissant ces calculs sont issues de l'expertise du domaine. Les calculs de tels indicateurs peuvent s'inscrire dans une méthode donnant les calculs et les agrégations de résultats de bas niveau en résultats de plus haut niveau compréhensibles par l'expert effectuant l'évaluation. Voici quelques exemples : la méthode Indigo [INRA, 2002], le logiciel de gestion des Déchets de Produits de Protection des Plantes Diaphyt [ACTA, ICTA, INRA , 2004], Dialogue [Solagro, 2002], Dialecte [Solagro, 1999], Diage [FRCA Centre, 2004], le diagnostic de durabilité du RAD [RAD, 2010], le diagnostic PRAIRIE [RAD, 2009], DAEG [Ossard et al., 2009], MASC (Multi-attribute assessment of the Sustainability of Cropping systems) [Sadok et al., 2009], IDEA (Indicateurs de Durabilité des Exploitations Agricoles) [Vilain et al., 2003]. Ces modèles sont pour la plupart multicritères i.e. évaluent les performances des agrosystèmes par rapport à des critères plus larges que les performances agronomiques en terme de rendement uniquement. Les critères élargis d'évaluation des systèmes de culture comprennent entre autres les indicateurs d'impacts environnementaux tels que les ressources et la qualité des sols, la qualité de l'eau et de l'air, la biodiversité et le bilan énergétique d'une production. Les objectifs d'amélioration des systèmes de cultures ainsi que leur évaluation peuvent être élargis aux domaines économique, social et environnemental grâce aux outils tels que MASC ou IDEA, que nous présenterons ci-après.

---

<sup>18</sup> Outil d'Aide à la Décision

### 3.1.8.1 Évaluation multicritère de la contribution des systèmes de culture au développement durable (indicateur MASC)

L'indicateur MASC (Multi-attribute assessment of the Sustainability of Cropping systems) est modélisé sous forme d'arbre dont les feuilles sont composées d'indicateurs socio-économiques, environnementaux et territoriaux tels que le chiffre d'affaires, la surface cultivée, le nombre d'employés par unité de surface, le niveau des intrants, etc. L'indice est donc composite, avec des poids différents en fonctions des « sous-indicateurs » qui ont été définis par les chercheurs de l'INRA qui ont développé ce modèle [Sadok et al., 2009]. Concrètement, l'application pour l'évaluation d'une exploitation agricole par le modèle MASC se base sur le logiciel d'aide à la décision multi-attributs nommé « DEXi » [Bohanec et al., 2013], qui permet l'évaluation interactive de modèles d'aide à la décision multicritères qualitatifs complexes. Ce modèle multi-attributs est une structure hiérarchique qui décompose le problème en sous-problèmes individuellement moins complexes et par conséquent plus faciles à résoudre que le problème complet. Le modèle MASC version 2.0 ainsi que DEXi sont téléchargeables librement sur Internet<sup>19,20</sup>.

Les indicateurs de plus bas niveau correspondent à des données brutes transformées en indicateurs qualitatifs tels que « bon », « moyen », « mauvais ». Il s'agit en fait d'un cas particulier des opérateurs linguistiques (linguistic predicates) [Zadeh, 1975] qui consistent à raisonner sur des concepts linguistiques flous (imprécis). Ensuite, ces indicateurs qualitatifs sont agrégés à l'aide d'un tableau renseigné à dire d'experts par le biais d'un raisonnement qualitatif. Par exemple, un expert peut décider que pour un nœud donné l'agrégation de deux critères « élevés » donne pour résultat un critère « élevé », etc. Il n'y a donc pas de loi bien précise pour agréger de tels indicateurs. Pour automatiser le processus d'agrégation de ces « linguistic predicates », on pourrait utiliser les opérateurs classiques de logique floue, à savoir des T-normes et T-conormes (par exemple les opérateurs *min* et *max*).

Un exemple d'arborescence hiérarchique pour l'évaluation de la contribution d'une exploitation au développement durable dans le modèle MASC 2.0 est représenté dans l'illustration suivante :

---

<sup>19</sup> Téléchargement de MASC : <http://wiki.inra.fr/wiki/deximasc/package+MASC/WebHome>

<sup>20</sup> Téléchargement de DEXi : <http://kt.ijs.si/MarkoBohanec/dexi.html>

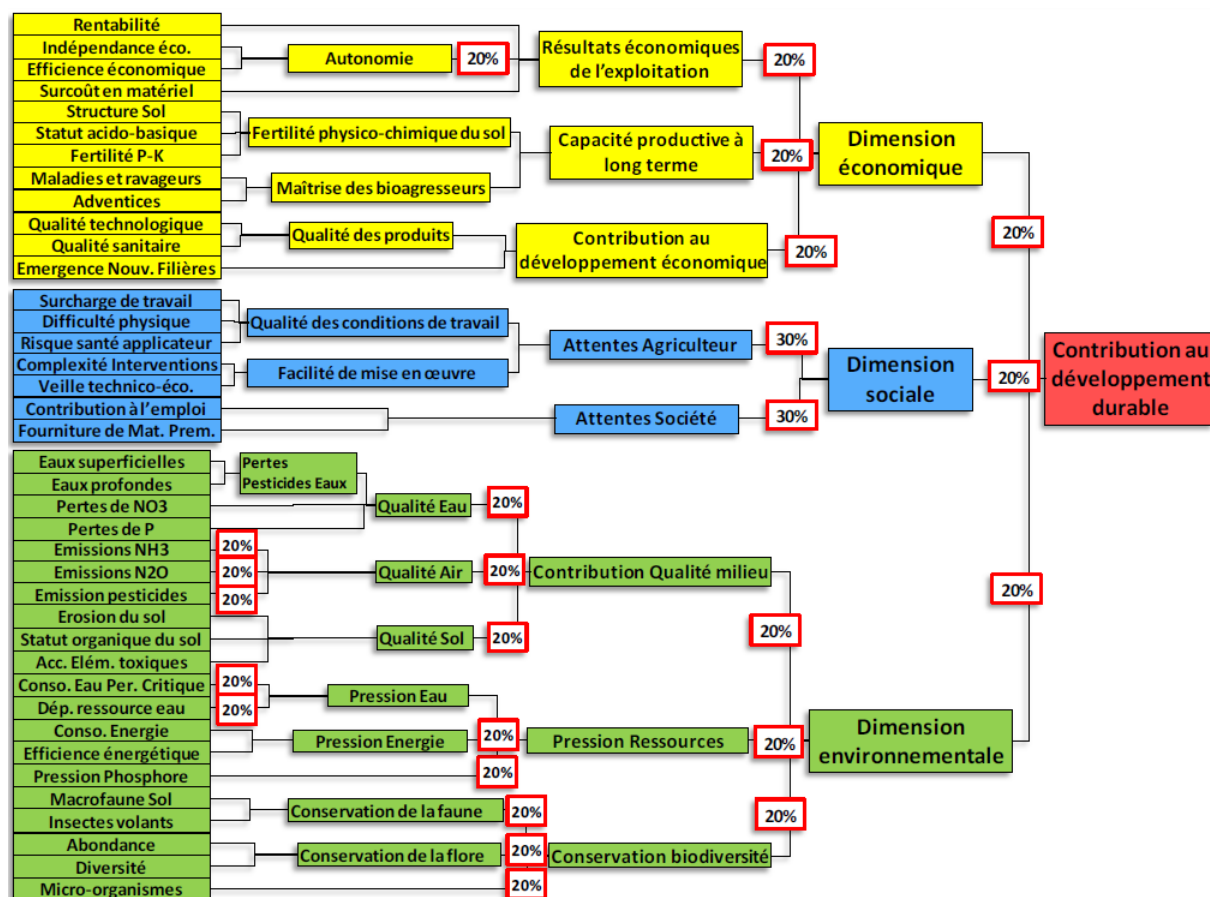


Figure 9 : arborescence hiérarchique de la contribution au développement durable dans le modèle MASC 2.0 (source : [Craheix et al., 2012])

### 3.1.8.2 Évaluation de la durabilité des exploitations agricoles (indicateur IDEA)

IDEA (Indicateurs de Durabilité des Exploitations Agricoles) [Vilain et al., 2003] est un outil à vocation pédagogique qui permet de sensibiliser les étudiants au concept de durabilité en agriculture. Cet indicateur évalue la durabilité selon 3 échelles : l'échelle de durabilité agro-écologique, l'échelle de durabilité socio-territoriale, l'échelle de durabilité économique. Pour chacune de ces échelles de durabilité, les scores obtenus par quantification des caractéristiques de l'exploitation prédéfinis sont agrégés. Le modèle utilisé comprend dix-sept objectifs agroécologiques qui doivent permettre une bonne efficacité économique pour un coût écologique aussi faible que possible, tels que l'autonomie, la biodiversité, l'organisation de l'espace, la protection des paysages, le bien-être animal, la qualité des produits, l'éthique, la qualité de vie, l'emploi, etc.



Pour passer du cadre conceptuel des dix-sept objectifs au calcul de durabilité, les objectifs sont traduits en critères mesurables appelés indicateurs. Dans cette phase, il est utile de construire une matrice croisant les objectifs avec les indicateurs.

		Objectifs																	
		N° de l'indicateur	Coherence	Autonomie	Biodiversité	Protection paysages	Protection des sols	Protection et gestion de l'eau	Air atmosphère	Ressources non renouvelables	Bien-être animal	Qualité des produits	Ethique	Développement humain	Développement local	Qualité de vie	Citoyenneté	Adaptabilité	Emploi
10 composantes et 42 indicateurs	Diversité	A1	1	1	1	1	1			1									
		A2	1	1	1	1	1			1									
		A3	1	1	1														
		A4	1		1														
	Organisation de l'espace	A5	1	1	1	1	1	1	1										
		A6	1		1	1	1	1	1										
		A7	1	1	1	1		1											
		A8	1	1	1	1	1	1	1		1								
		A9	1		1	1	1										1		
		A10	1	1	1	1	1	1			1	1							
		A11	1	1	1	1	1	1	1			1							
	Pratiques agricoles	A12	1	1	1				1	1	1	1							
		A13							1	1	1	1					1		
		A14	1	1	1			1	1	1		1							
		A15	1	1	1			1			1	1					1		
		A16			1			1	1		1								
		A17		1	1			1	1		1						1		
		A18	1	1					1	1									
	Qualité des produits et des territoires	B1					1				1	1		1	1		1		1
		B2	1											1	1				
		B3					1		1		1						1	1	
		B4	1										1				1	1	
		B5	1										1	1	1	1	1	1	
	Emploi et services	B6	1	1						1						1			
		B7	1	1				1		1						1			
		B8														1			
		B9											1			1	1		1
		B10												1	1	1	1		
		B11													1	1			1
	Ethique et développement humain	B12	1	1						1			1	1					
		B13									1	1	1				1		
		B14	1											1	1	1		1	1
		B15	1											1		1			1
		B16											1	1		1			
		B17												1		1			
		B18						1					1	1		1	1		
	Viabilité	C1	1													1	1		1
		C2	1															1	
	Indépendance	C3	1	1													1		1
		C4	1	1														1	
	Transmissibilité	C5	1																
Efficience	C6	1	1						1						1	1		1	

Figure 10 : exemple de matrice croisant les objectifs avec les indicateurs (source : [Vilain et al., 2003])

Chaque échelle de durabilité est subdivisée en trois ou quatre composantes qui synthétisent les caractéristiques fondamentales du diagnostic de durabilité. Par exemple, la durabilité agro-écologique est subdivisée en trois composantes : la diversité, l'organisation de l'espace et les pratiques agricoles. Ces composantes sont à leur tour subdivisées en indicateurs.

Le mode de calcul est basé sur un système de points avec un plafonnement ; les trois échelles de durabilité, de même poids, varient de 0 à 100 points. Le score d'une exploitation pour chacune des trois échelles de durabilité est le nombre cumulé de points de durabilité obtenus pour les indicateurs de l'échelle considérée. Plus la note est élevée, plus l'exploitation est considérée comme durable pour l'échelle considérée. Chaque composante est, de la même manière, plafonnée à environ 33 points. Ceci implique que de très nombreuses combinaisons

permettent d'atteindre un même degré de durabilité. En effet, même si certains principes sont communs à tous les systèmes agricoles, les concepteurs d'IDEA considèrent qu'il n'y a pas un système agricole « idéal » unique.

Composantes		Indicateurs	Valeurs maximales	
<b>Diversité domestique</b>	A1	Diversité des cultures annuelles ou temporaires	14	Total plafonné à 33 unités
	A2	Diversité des cultures pérennes	14	
	A3	Diversité animale	14	
	A4	Valorisation et conservation du patrimoine génétique	6	
<b>Organisation de l'espace</b>	A5	Assolement	8	Total plafonné à 33 unités
	A6	Dimension des parcelles	6	
	A7	Gestion des matières organiques	5	
	A8	Zones de régulation écologique	12	
	A9	Contribution aux enjeux environnementaux du territoire	4	
	A10	Valorisation de l'espace	5	
<b>Pratiques agricoles</b>	A11	Gestion des surfaces fourragères	3	Total plafonné à 34 unités
	A12	Fertilisation	8	
	A13	Effluents organiques liquides	3	
	A14	Pesticides	13	
	A15	Traitements vétérinaires	3	
	A16	Protection de la ressource sol	5	
	A17	Gestion de la ressource en eau	4	
	A18	Dépendance énergétique	10	

Figure 11 : échelle de durabilité agro-écologique subdivisée en indicateurs (source : [Vilain et al., 2003])

Les barèmes de notation et pondérations associés à chaque indicateur sont des fonctions d'utilité qui sont réalisées par des experts des domaines en jeu. Le résultat obtenu pour chaque indicateur (par exemple, la somme pondérée des doses des différents pesticides épandus sur chaque parcelle - indicateur « A14 – pesticides ») est discrétisé et est inversement proportionnel en nombre de points affectés à l'indicateur. Le système d'agrégation des indicateurs est l'addition des points obtenus par les différents indicateurs, ce qui sous-entend une compensation entre critères favorables et défavorables pour les différentes composantes.

### 3.1.9 Conclusion et limites des approches d'aide à la décision

Lors de leurs recherches sur la théorie de la décision, les chercheurs ont, de tout temps, essayé d'établir des méthodes rigoureuses mathématiquement afin que le décideur puisse prendre les meilleures décisions possibles. Ces méthodes se sont multipliées par le fait que des décideurs peuvent avoir des objectifs différents face à une même situation. Chaque décideur possède aussi sa sensibilité, son aversion au risque, etc. Outre l'adaptation au décideur, la méthode à utiliser et son résultat dépend grandement de puissance de calcul dont on dispose mais aussi

des caractéristiques des informations en entrée, en termes de quantité, de qualités, de distribution, etc. Toute la problématique est de choisir la méthode qui convient au contexte et au décideur : c'est ce qui permettra de prendre la meilleure décision.

L'inconvénient majeur des applications présentées est leur multiplicité due à leur spécialisation. En effet chaque indicateur propose des sous-indicateurs ainsi que des méthodes d'agrégations différentes, ce qui peut entraîner des diagnostics différents pour une même exploitation, en fonction de l'indicateur utilisé. Plutôt que de proposer une méthode unique, nous proposerons de réaliser un méta-modèle générique de gestion de l'imperfection des données ainsi que des opérateurs d'agrégation. Il apparaît en effet nécessaire de proposer un cadre formel rigoureux par exemple pour l'agrégation d'indicateurs, afin d'atténuer le caractère subjectif de l'évaluation de la durabilité avec de tels modèles et d'automatiser les calculs. Enfin, l'agrégation de tels indicateurs nécessite la prise en compte des différents types d'imperfection susceptibles d'entacher les données. Ces apports permettront au décideur de choisir la meilleure méthode en vue de prendre la meilleure décision.

Dans le prochain paragraphe, nous aborderons les approches mathématiques de modélisation et de manipulation d'informations imparfaites.

## 3.2 Modélisation et manipulation d'informations imparfaites

### 3.2.1 Introduction

D'après [Dubois et Prade, 1988], on peut décomposer l'imperfection des données en différentes catégories non exclusives : l'incertitude, l'imprécision et l'inconsistance (incohérence), qui peuvent elles-mêmes se décomposer en sous-catégories. L'incertitude d'une information peut faire référence à une variable qualitative dont plusieurs valeurs sont possibles ou bien à des données quantitatives appartenant à un intervalle. Elle fait également référence à la véracité d'une information. Par exemple, une information peut se révéler complète, précise mais fausse [Masson, 2005]. Les différents types d'imperfection des données sont illustrés par la figure suivante :

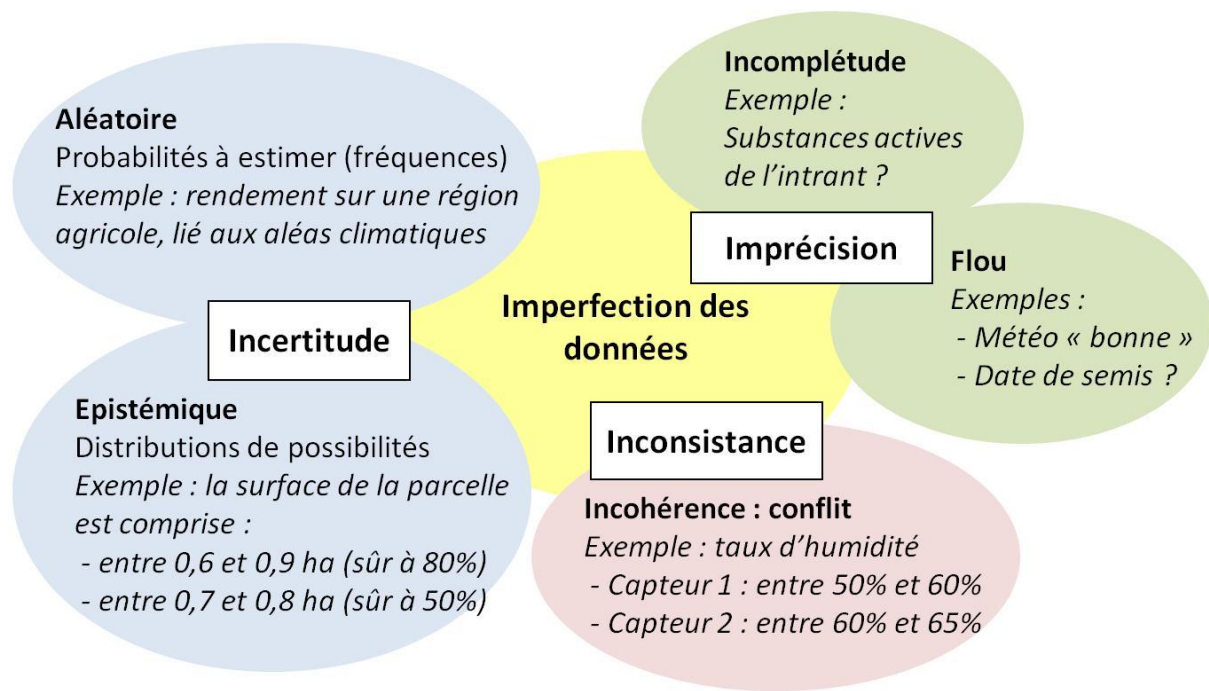


Figure 12 : différents types d'imperfection des données

(1) On peut distinguer deux types d'incertitude [Zadeh, 1978], [Dubois et Prade, 1988] :

- L'incertitude objective que l'on peut assimiler à un aléa (la mesure est alors soumise à une certaine variabilité) que nous allons appeler incertitude probabiliste ou aléatoire. Dans ce cadre, l'information est d'autant plus précise et certaine que le nombre d'échantillons mesurés est grand. Cependant, réduire l'incertitude d'une donnée à son caractère aléatoire n'est pas satisfaisant.
- L'incertitude subjective liée au manque de données ou de connaissances, qui peut être aussi liée au crédit que l'on accorde à la source d'information, que nous allons appeler incertitude possibiliste ou épistémique. Dans la théorie des possibilités, on met en évidence l'antagonisme entre imprécision et incertitude. Ceci est illustré par le fait qu'augmenter la précision du contenu d'une proposition augmente son degré d'incertitude. Inversement, le caractère incertain d'une information précise a tendance à accroître l'imprécision des conclusions qu'on pourrait tirer d'une situation.

(2) On peut aussi modéliser l'imprécision qui fait référence soit à l'incomplétude (absence d'information) soit au caractère vague (flou) d'une information, qui peut être subjectif [Zadeh, 1965].

(3) Enfin, l'inconsistance (incohérence) survient en présence de redondance et lorsque plusieurs informations sont en conflit. Elle peut notamment être traitée par une fusion de données [Pivert et Prade, 2012].

### 3.2.2 L'analyse de l'imperfection par les mesures de confiance

De tout temps, les chercheurs ont manifesté un intérêt croissant pour les différentes méthodes d'analyse de l'imperfection. Ces approches s'inscrivent dans le cadre général des mesures de confiance. Dans les prochains paragraphes, nous allons donc présenter différentes approches élaborées en vue de mesurer la confiance que l'on peut accorder à un événement en jeu dans les systèmes socio-environnementaux. En effet, la théorie des probabilités s'inscrit dans un éventail d'autres théories comme la théorie des possibilités ou la théorie de Dempster-Shafer. Ces théories sont pour la plupart moins restrictives et permettent notamment de modéliser l'incomplétude et l'imprécision relatives à des événements, que la théorie des probabilités ne peut modéliser efficacement. D'une manière générale, la confiance dans un événement est définie sur un ensemble  $\Omega$  d'événements. Nous parlerons d'ensembles finis auquel cas les événements sont tous des parties de  $\Omega$  ; mais au cas où  $\Omega$  n'est pas fini, les événements doivent former une tribu de parties, c'est à dire un ensemble de parties comprenant  $\Omega$ , stable par complémentation, intersections et unions finies ou dénombrables. On nomme alors une mesure de confiance, toute fonction  $c$  d'une tribu de  $\Omega$  vers  $[0, 1]$  vérifiant les propriétés suivantes : (1)  $c(\emptyset) = 0$  et  $c(\Omega) = 1$  ; (2)  $A \subset B \Rightarrow c(A) \leq c(B)$  ; (3) pour toute suite croissante ou décroissante :  $A_1 \subset A_2 \subset A_3 \subset \dots \subset A_n \Rightarrow \lim c(A_n) = c(\lim(A_n))$  (continuité) [Gacogne, 1997]. En conséquence on aura :  $c(A \cup B) \geq \max(c(A), c(B))$  et  $c(A \cap B) \leq \min(c(A), c(B))$ . La mesure de confiance la plus répandue est la théorie des probabilités, qui a longtemps été la principale méthode de modélisation de l'incertitude. Si la théorie des probabilités est toujours largement utilisée, d'autres méthodes sont apparues durant ces quarante dernières années. La théorie des possibilités modélise le degré d'appartenance d'un événement à un ensemble, qui est mesuré dans le cas le plus favorable, c'est-à-dire que la possibilité d'une disjonction d'événements est égale au maximum des possibilités d'occurrence de ces événements. Une mesure de possibilité possède une mesure duale de nécessité. Ces représentations sont davantage adaptées à la représentation de l'ignorance totale, ainsi qu'aux événements qui ne sont pas liés à des phénomènes aléatoires et reproductibles. Ensuite, nous pouvons citer la théorie de Dempster-Shafer, qui modélise les

degrés d'évidence contenue par une partie de  $\Omega$  (crédibilité), ainsi que l'évidence que peut potentiellement générer un événement (plausibilité).

### 3.2.3 Modélisation de phénomènes aléatoires

Pendant très longtemps la modélisation de l'incertitude est restée abordée par la théorie des probabilités, étude mathématique des phénomènes caractérisés par le hasard et l'incertitude. La probabilité d'un événement est égale à la somme des probabilités d'occurrence de ces événements, si, par hypothèse, ceux-ci sont disjoints deux-à-deux. Les distributions de probabilité caractérisent en général des phénomènes aléatoires : elles décrivent soit des probabilités d'événements modélisées par une variable aléatoire discrète, soit des probabilités d'appartenance à un intervalle modélisées par une variable aléatoire continue. La statistique consiste à collecter, traiter et interpréter des ensembles de données issues de la réalité. Grâce aux distributions de probabilité, la variabilité naturelle des phénomènes aléatoires peut être modélisée, par exemple la concentration de la pollution d'un sol, la production agricole d'une parcelle au fur et à mesure des années, etc. Dans cette approche, on associe aux événements d'un référentiel  $\Omega$  (ensemble fini d'événements), une mesure de probabilité, c'est à dire une application  $P$  d'un  $\sigma$ -algèbre  $P(\Omega)$ , ensemble des parties de  $\Omega$ , dans  $[0,1]$ , vérifiant  $P(\Omega) = 1$ . De plus,  $P$  est  $\sigma$ -additive, i.e.  $\forall (A_i)_{i=1}^{i=n \in \mathbb{N}^*} \in P(\Omega)$ ,  $P(\cup_i A_i) = \sum_i P(A_i)$  (contrainte d'additivité), où les  $\{A_i\}_{i \in \{1, \dots, n\}}$  forment une famille dénombrable d'événements disjoints deux à deux de  $P(\Omega)$ . Ces axiomes impliquent notamment que la probabilité d'un événement détermine complètement celle de son contraire, i.e.  $P(A) = 1 - P(\overline{A})$ . Cependant, cette approche est peu adaptée à la représentation de l'ignorance totale. De plus, l'interprétation objective des probabilités affectées aux événements se révèle difficile lorsque les connaissances manipulées ne sont plus liées à des phénomènes aléatoires et reproductibles [Dubois et Prade, 1988] ou si elles sont liées à des nombres d'observations trop faibles. En effet, lorsqu'une trop grande imprécision des informations ne permet pas une évaluation suffisante de l'incertitude, l'approche classique qui consiste à modéliser les événements par une loi de probabilité uniforme d'étendue égale à la largeur de l'intervalle d'incertitude (par exemple lorsque l'état de la connaissance ne permet que de juger que le paramètre incertain  $x$  se trouve dans l'intervalle  $[x - \Delta x, x + \Delta x]$ ) n'est pas satisfaisante car elle modélise en général une information plus riche que la connaissance que l'on a du monde réel (probabilités trop « précises ») et peut en modéliser plusieurs états différents : nous risquons alors de sous-

estimer ou de surestimer le résultat final. Les travaux de [Ferson et Ginzburg, 1996] préconisent de rechercher d'autres outils mathématiques plus flexibles selon la donnée effectivement disponible pour traiter l'incertitude épistémique. Par exemple, le cas limite de l'ignorance totale est très mal modélisé par la théorie des probabilités. Une modélisation probabiliste suppose qu'on ait identifié un certain nombre d'événements disjoints auxquels on affecte une probabilité identique qui est l'inverse de ce nombre d'événements selon le principe d'entropie maximale (équiprobabilité).

### 3.2.4 Modélisation d'informations incertaines, incomplètes, imprécises, linguistiques

#### 3.2.4.1 La théorie des sous ensembles flous

La théorie des sous ensembles flous (« fuzzy sets ») a été initialement introduite par L. Zadeh [Zadeh, 1965]. Il s'agit d'une extension de la théorie des ensembles classiques aux ensembles définis de façon imprécise. Cette théorie sert de fondement à la logique floue. Ainsi un sous-ensemble flou  $F$  d'un ensemble  $E$ , se définit au moyen d'une fonction d'appartenance  $\mu_F$  de  $E$  dans  $[0,1]$ , qui associe à tout élément  $x$  de  $E$  son degré d'appartenance  $\mu_F(x)$  au sous-ensemble  $F$  (c'est-à-dire que  $x$  appartient « plus ou moins » à  $F$ ). Lorsque cette fonction d'appartenance est normalisée (il existe une valeur  $x$  de  $E$  telle que  $\mu_F(x) = 1$ ) on peut alors interpréter  $\mu_F(x)$  comme la possibilité que  $F$  prenne la valeur  $x$  ( $\mu_F(x)$  est alors une distribution de possibilité). Nous parlons alors de sous-ensemble flou de type 1, que nous noterons par la suite T1FS<sup>21</sup>. En pratique, on peut associer des variables symboliques (e.g. des mots du langage courant tels que « petit », « grand », etc.) à des sous-ensembles flous (on parle alors de « variables linguistiques »), permettant ainsi d'automatiser des raisonnements lors de la mise en œuvre de systèmes flous.

Les sous-ensembles flous de type 2, que nous noterons par la suite T2FS<sup>22</sup>, ont été introduits par l'argument que les sous-ensembles flous de type 1 modélisent uniquement le caractère vague de l'information, mais pas l'incertitude. Afin de traiter l'incertitude, [Zadeh, 1975] a introduit les « generalized » T2FS (sous-ensembles flous généralisés de type 2), dont la

---

<sup>21</sup> T1FS est en réalité l'abréviation de « Type-1 Fuzzy Set »

<sup>22</sup> T2FS est en réalité l'abréviation de « Type-2 Fuzzy Set »

fonction d'appartenance est tridimensionnelle. La troisième dimension de cet ensemble est la valeur de la fonction d'appartenance en chaque point du domaine à deux dimensions qui est appelé la FOU<sup>23</sup>. La FOU représente le caractère flou d'une fonction d'appartenance de type 1, et est caractérisée complètement par ses deux fonctions englobantes ; une fonction d'appartenance basse (lower membership function ou LMF) et une fonction d'appartenance haute (upper membership function ou UMF), qui représentent chacune des T1FS [Mendel et John, 2002]. Un cas particulier des « generalized T2FS » est les « interval T2FS » (sous-ensembles flous intervalles de type 2), dont les fonctions d'appartenance sont des intervalles nets. Ils ont reçu une attention particulière dans la littérature en raison de leur complexité de calcul moindre [Coupland et John, 2008], y compris dans le domaine de l'aide à la décision [Gong et al., 2015].

#### 3.2.4.2 Les L-fuzzy-sets

Dans la théorie des sous-ensembles flous, le degré d'appartenance d'un sous-ensemble flou est défini sur  $[0, 1]$ . Goguen a montré que ce dernier pouvait être à valeurs dans un treillis : c'est la théorie des L-fuzzy-sets<sup>24</sup> [Goguen, 1967]. Un treillis (« lattice ») est un ensemble partiellement ordonné dans lequel chaque couple d'éléments admet une borne supérieure et une borne inférieure. Il est muni d'une relation d'ordre  $\subseteq$  réflexive, transitive et antisymétrique. La borne supérieure et la borne inférieure définissent alors deux lois internes  $\wedge$  et  $\vee$  commutatives, associatives et idempotentes ;  $\forall (a, b) \in E^2, \inf(a, b) = a \wedge b$  et  $\sup(a, b) = a \vee b$ . On définit donc un sous-ensemble L-flou comme étant une application de  $E$  dans  $L$ . Si  $L = [0, 1]$ , on retrouve la théorie « classique » des sous-ensembles flous introduite initialement par Zadeh. Les L-fuzzy sets permettent de considérer le flou intrinsèque aux attributs ainsi que l'indétermination qui se rattache aux degrés de vérité.

#### 3.2.4.3 Modélisation prudente des incertitudes : l'arithmétique d'intervalles

L'arithmétique d'intervalles a été, dans un premier temps, utilisée pour contrôler les erreurs d'arrondis causées par le calcul sur des nombres à virgule flottante sur ordinateur [Moore, 1966] [Moore, 1979]. En effet, l'ensemble des réels étant infini et indénombrable, les résultats

---

<sup>23</sup> « footprint of uncertainty », traduit littéralement par « empreinte d'incertitude »

<sup>24</sup> Le « L » de L-fuzzy-set signifie « Lattice » (treillis)



de calculs numériques sur des nombres flottants peuvent engendrer des erreurs significatives lors de leur traitement sur ordinateur, considérant que leur représentation en mémoire est par définition limitée. Les nombres réels doivent en circonstance être représentés par des approximations, c'est-à-dire des intervalles sur lesquels sont effectués les calculs. La méconnaissance de la valeur exacte d'une variable nécessite de modéliser celle-ci par un intervalle. C'est également le cas lorsque des variables représentent des mesures de quantités physiques qui sont réalisées avec une certaine imprécision due aux appareils de mesure (e.g. bornes de l'erreur<sup>25</sup> sur une mesure). On appelle cette approche arithmétique d'intervalles, calcul par intervalle, ou encore « calcul d'erreurs » en physique.

On utilise un couple de nombres pour représenter ces intervalles d'incertitude ; par exemple l'intervalle  $x = [x_{\text{inf}}, x_{\text{sup}}]$  ou bien  $x = [x - \delta x, x + \delta x]$  qui définit un intervalle autour de la valeur mesurée qui inclut la valeur vraie avec une probabilité plus ou moins grande en cas d'absence d'erreur systématique.

Dans le cas d'arrondi de nombres flottants, l'addition est définie comme étant l'intervalle  $[x_{\text{inf}} +_{\text{inf}} y_{\text{inf}}, x_{\text{sup}} +_{\text{sup}} y_{\text{sup}}]$ .  $+_{\text{inf}}$  (respectivement  $+_{\text{sup}}$ ) correspond à l'opérateur « addition » de deux nombres dont le résultat est arrondi au nombre flottant « représentable informatiquement » le plus proche dans la direction  $-\infty$  (respectivement  $+\infty$ ). Ces arrondis permettent de conserver la propriété d'inclusion : pour tout couple de valeurs réelles dans les intervalles  $x$  et  $y$ , leur somme se trouve nécessairement dans l'intervalle résultat  $x + y$ . Dans le cas de mesures physiques entachées d'erreurs aléatoires, on cherche généralement à évaluer le degré de l'erreur aléatoire commise, soit en trouvant un majorant (ce cas conduit à des intervalles très larges au fur et à mesure des calculs) soit par l'analyse statistique d'un ensemble de mesures permettant de caractériser leur distribution de probabilité des valeurs prises et d'estimer la valeur moyenne et la variance de cette distribution.

On peut alors effectuer des opérations arithmétiques sur les différents intervalles, comme sur les nombres réels, telle que l'addition de deux intervalles  $x + y$ . Par exemple, l'addition des

---

<sup>25</sup> Dans le cadre de mesures physiques on distingue en général les erreurs aléatoires, caractérisées par une distribution de probabilité répartie autour de la valeur vraie dans le cas d'erreurs purement aléatoires (on parle alors d'échantillon Bernoullien), des erreurs systématiques qu'une étude statistique ne peut détecter, comme par exemple une erreur d'étalonnage (valeur vraie inconnue)

intervalles  $[-10, -6]$  et  $[8, 10]$  donne le résultat  $[-2, 4]$  qui contient 0 (signe indéterminé). De manière similaire, on définit d'autres opérateurs arithmétiques tels que la soustraction, la multiplication, la division, les puissances, etc. L'ensemble des intervalles, contrairement aux nombres réels, n'admet pas de relation d'ordre total pour les opérateurs de comparaison. Cette méthode permet donc de propager ces intervalles par des calculs simples pour les opérations arithmétiques de base telles que  $+$ ,  $-$ ,  $*$  et  $/$  (sous réserve de prendre en compte des cas particuliers tels que la présence de 0 dans un intervalle dans le cas d'une division). Il est également possible d'approximer des calculs via une différentiation (développement en série de Taylor) à l'ordre 1, notamment dans le cas d'une fonction non linéaire s'appliquant à un intervalle. En pratique l'arithmétique d'intervalles permet de se focaliser sur les calculs des enveloppes basse et haute, ce qui réduit le nombre de calculs à effectuer. Dans ce cas, au lieu de  $n$  ( $n$  très grand) calculs avec une méthode de simulation de Monte Carlo, il n'y a besoin que de résoudre deux équations pour les enveloppes basse et haute, pour chaque inconnue [Guillard et al., 2011]. D'un point de vue « coût computationnel », cette méthode est donc bien plus efficace que les méthodes probabilistes.

Cependant, par définition, cette approche se prête mal aux événements répétitifs dans le temps pour lesquels une modélisation probabiliste rigoureuse est plus adaptée. De plus, elle est trop « prudente » et manque de précision. En effet, si l'on enchaîne de nombreux calculs, on peut parfois obtenir des cas où « presque tout devient possible », en raison de son caractère peu spécifique. Enfin, l'arithmétique d'intervalles est un cas particulier de l'approche floue en considérant un noyau identique au support.

#### 3.2.4.4 La théorie des possibilités

La théorie des possibilités [Dubois et Prade, 1988] [Zadeh, 1978] est un cadre formel pour la prise en compte de l'imprécision (degré de vérité des sous-ensembles flous) et de l'incertitude (vérité ou fausseté d'une proposition) dans des connaissances, ainsi que du raisonnement sur de telles connaissances. En effet, cette théorie permet d'exprimer dans quelle mesure la réalisation d'un événement est possible et dans quelle mesure on en est certain, sans toutefois avoir à disposition la probabilité de cette réalisation. Une distribution de possibilité est un ensemble flou normalisé de valeurs mutuellement exclusives. La théorie des possibilités supprime la contrainte forte d'additivité probabiliste et associe aux événements du référentiel  $\Omega$  une mesure de possibilité (degré de croyance) notée  $\Pi$ , et une mesure de nécessité (degré de certitude) notée  $N$ , vérifiant les relations suivantes :  $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$  et

$N(A \cap B) = \min(N(A), N(B))$ . La relation reliant la possibilité d'un événement et son contraire est alors donnée par  $\Pi(A) = 1 - N(\bar{A})$  et l'ignorance totale se traduit alors par l'équation  $\Pi(A) = \Pi(\bar{A}) = 1$ , ce qui implique:  $N(A) = N(\bar{A}) = 0$ . La théorie des possibilités est une représentation unificatrice d'informations incertaines, incomplètes, imprécises et linguistiques.

Dans la théorie des possibilités, l'ignorance n'est alors pas modélisée par l'équiprobabilité mais par une possibilité égale à 1 sur tous les événements possibles. Par exemple, pour  $Y = X^2$ , par le principe d'extension de Zadeh, l'ignorance sur  $X$  conduit bien à l'ignorance sur  $Y$  (alors que le carré d'une variable aléatoire uniforme conduit à une variable aléatoire triangulaire). En effet, la possibilité est reliée à la distribution et non pas à la densité. La distribution maximale spécifique existe toujours et devient de plus en plus spécifique quand on ajoute de l'information (par exemple unimodalité). Ainsi, le modèle probabiliste est adapté au traitement d'informations précises mais dispersées. La théorie des possibilités permet en revanche de modéliser des connaissances imprécises mais cohérentes [Dubois et Prade, 1988].

### 3.2.5 Modélisation de connaissances à la fois aléatoires et floues : les « fuzzy random variables »

Les interactions entre les variables floues et les variables aléatoires, et plus particulièrement leur complémentarité, ont été largement explorées dans la littérature, notamment sous le nom de « fuzzy random variables » (FRVs, traduit littéralement par les variables aléatoires floues). Trois principales définitions de FRVs ont été étudiées : [Kwakernaak, 1978] et [Kwakernaak, 1979] qui ont inventé le concept de FRV et [Kruse et Meyer, 1987] qui les ont définies comme étant des variables aléatoires, dont les valeurs ne sont pas réelles, mais des nombres flous. (2) [Puri et Ralescu, 1986] et [Klement et al., 1986] ont défini les FRVs comme des sous-ensembles flous aléatoires, c'est-à-dire la fuzzification d'un ensemble aléatoire. Leur approche a permis de palier deux limites du modèle de Kwakernaak : son mapping sur les nombres réels plutôt que sur un espace euclidien de dimension  $n$  et sa notion de mesurabilité qui est limitée aux nombres flous réels. Enfin, (3) [Liu et Liu, 2003] ont basé la notion de FRV sur un concept appelé mesure de crédibilité. Une synthèse ainsi que des exemples sur les différentes approches relatives aux FRVs sont présentes dans [Shapiro, 2012]. En résumé, les trois approches peuvent être différenciées de par la forme (i.e. variable floue ou scalaire) de l'espérance et de la variance, comme l'illustre le tableau suivant :

Tableau 7 : types d'incertitudes de l'espérance et de la variance pour les différentes FRVs [Shapiro, 2012]

Auteur(s)	Espérance	Variance
Kwakernaak	Floue	Floue
Puri et Ralescu	Floue	Scalaire
Liu et Liu	Scalaire	Scalaire

D'un point de vue pratique, l'idée générale est la caractérisation d'une suite d'observations floues dont la réalité est précise et est une variable aléatoire. Par exemple, supposons que l'on attribue de manière approximative la productivité de chaque parcelle lorsque l'on parcourt une exploitation. La productivité réelle  $X$  est une variable aléatoire réelle. Toutefois, la précision de notre perception ne nous permet que d'avoir accès à une variable aléatoire  $X'$  discrète dont les valeurs sont par exemple « catastrophique, faible, moyen, élevé, exceptionnel ». Nous percevons donc des observations floues tandis que la variable aléatoire  $X$  réelle n'est pas observable.

### 3.2.6 Combinaison d'informations incertaines, incomplètes, imprécises, linguistiques et aléatoires : la théorie de Dempster-Shafer

Les travaux de G. Choquet [Choquet, 1953] et A. Dempster [Dempster, 1967] sur les bornes inférieure et supérieure d'une famille de distributions de probabilité ont permis à G. Shafer d'établir les bases de la théorie de l'évidence. Ce formalisme théorique a été ensuite utilisé pour développer la théorie des fonctions de croyance de Shafer et a montré son intérêt pour modéliser les connaissances incertaines [Shafer, 1976]. Ces théories sont communément regroupées sous le nom de théorie de Dempster-Shafer. Ensuite, Smets [Smets et Kennes, 1994] a développé le Modèle des Croyances Transférables (MCT ; traduction de « Transferable Belief Model » ou « TBM »), un modèle non probabiliste de raisonnement sur l'incertain basé sur les fonctions de croyance, complétant ainsi la théorie de Dempster-Shafer et suscitant ensuite de nombreuses applications nœud [Denœux, 2010].

Soit  $A$  un sous-ensemble de l'ensemble fini  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}_{n \in \mathbb{N}}$  appelé cadre de discernement. Une fonction de croyance est définie par son « basic belief assignment » (ou « bba », traduit par affectation de masse de base ou jeu de masses)  $m$ , défini sur  $2^\Omega$  (i.e.

l'ensemble des parties de  $\Omega$  dans  $[0,1]$ , qui satisfait la relation  $\sum_{A \subseteq \Omega} m(A) = 1$ . Les valeurs de  $m(A)$  sont appelées « basic belief masses » (ou « bbm », traduit par masses de croyance de base). Chaque sous-ensemble  $A$  de  $\Omega$  avec  $m(A) > 0$  est appelé élément focal de  $m$ . La masse  $m(A)$  d'un sous-ensemble  $A \in 2^\Omega$  est le degré de croyance qui a été assigné spécifiquement à ce sous-ensemble, compte tenu de l'état actuel des connaissances. La masse  $m(A)$  est donc la proportion de toutes les preuves disponibles affirmant que l'état actuel est spécifiquement  $A$  (et non pas un autre état ou un sous-état de  $A$ ). La valeur de  $m(A)$  concerne donc uniquement le sous-ensemble  $A$  et n'apporte aucune preuve à ses propres sous-ensembles, chacun d'eux ayant, par définition, sa propre masse. L'ignorance totale est traduite par une masse vérifiant  $m(\Omega) = 1$  et  $m(A) = 0$  pour toute partie  $A$  non pleine de  $\Omega$ . Si par exemple, on ne considère que deux éventualités  $\Omega = \{a, b\}$  avec  $m(\{a\})$  et  $m(\{b\})$ , alors  $m(\{a, b\})$  symbolise l'indifférence à  $a$  ou  $b$ , et  $m(\varphi)$  symbolise la part de ce qui pourrait être autre chose que  $a$  ou  $b$ . On ne requiert donc pas  $m(\varphi) = 0$ , mais si c'est le cas,  $m$  est dite normale ou complète, et, alors, on peut montrer qu'une telle masse définit une famille de probabilités.

Le degré de croyance Bel de  $A$  évalue la quantité totale potentiellement affectée spécifiquement à  $A$ . Son calcul consiste à additionner toutes les masses allouées aux ensembles  $B$  qui sont inclus dans  $A$ , i.e.  $\forall A \subseteq \Omega, Bel(A) = \sum_{B \subseteq A, B \neq \varphi} m(B)$ . La mesure de plausibilité duale Pl de  $A$  évalue la quantité maximum d'évidence qui peut être potentiellement affectée à  $A$ . Elle est obtenue en additionnant toutes les masses allouées aux ensembles  $B$  qui n'ont pas d'intersection vide (et qui sont donc compatibles) avec  $A$ , i.e.  $\forall A \subseteq \Omega, Pl(A) = \sum_{B \cap A \neq \varphi} m(B)$ . Les mesures de crédibilité et de plausibilité vérifient la relation suivante :  $Pl(A) + Bel(\bar{A}) = 1 - m(\varphi)$ .

Lorsque la crédibilité et la plausibilité sont déduites d'une distribution de masse particulière que nous allons préciser ci-dessous, elles peuvent correspondre soit à une distribution de probabilité soit à une distribution de possibilité, comme le montrent les deux théorèmes qui suivent.

**Théorème : les éléments focaux sont totalement ordonnés par inclusion si et seulement si Bel et Pl sont respectivement une mesure de nécessité et de possibilité.**

Une distribution de possibilité n'est pas une fonction de croyance car, en général, la somme des possibilités est différente de 1. Une distribution de possibilité peut être transformée en distribution de masses focales de la manière suivante, dans le cas d'intervalles flous discrets :

soient  $I_1, I_2, \dots, I_i, \dots, I_n$   $n$  intervalles flous discrets. La relation entre les alpha-coupes  $\alpha_1 \geq \dots \geq \alpha_i \geq \dots \geq \alpha_n$  (avec  $\alpha_1 = 1$ ) caractérisant ces intervalles et les masses des éléments focaux associés  $m_1, \dots, m_i, \dots, m_n$  est la suivante :

$$m_1 = \alpha_1 - \alpha_2 ; m_2 = \alpha_2 - \alpha_3 ; \dots ; m_i = \alpha_i - \alpha_{i+1} ; \dots ; m_n = \alpha_n$$

En voici une illustration pour  $n = 4$  :

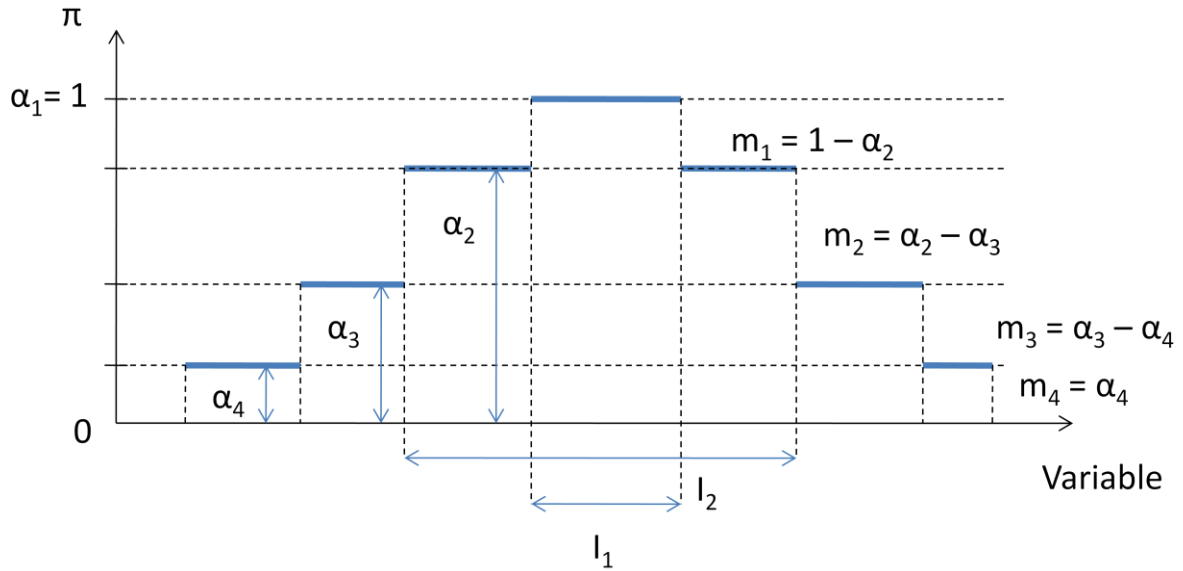


Figure 13 : relations entre alpha coupes et masses focales

**Théorème : une distribution de masse (croyance) est une probabilité si et seulement si les éléments focaux sont des singletons.**

Dans ce cas,  $P$  est une probabilité, c'est-à-dire une application de  $P(\Omega)$  vers  $[0,1]$ , satisfaisant les relations mathématiques suivantes :  $P(\Omega) = 1$  et  $P(\cup_i A_i) = \sum_i P(A_i)$  où les  $\{A_i\}_{i \in \{1, \dots, n\}}$  sont une famille finie d'événements disjoints de  $\Omega$ , et  $P(A) = 1 - P(\bar{A})$ .

Considérant deux sources d'informations, avec respectivement deux bba,  $m_1$  et  $m_2$ , définis sur  $\Omega$  ; ces deux jeux de masses peuvent être agrégés avec l'opérateur de combinaison conjonctive de Dempster (qui est aussi appelé somme orthogonale), qui correspond à l'opération de combinaison conjonctive de deux jeux de masses (qui est aussi appelé somme conjonctive) notée  $\cap$ , normalisé via un facteur qui dépend du conflit entre les deux jeux de masses :  $\left(\frac{1}{1 - m_\cap(\phi)}\right)$ .  $m_\cap(\phi)$  est le degré de conflit entre  $m_1$  et  $m_2$ . La distribution de bbm résultante a pour expression:

$$[\text{somme conjonctive}] \forall A \subseteq \Omega, m_{\cap}(A) = (m_1 \cap m_2)(A) = \sum_{B \cap C = A} m_1(B)m_2(C)$$

$$[\text{somme orthogonale}] \forall \phi \neq A \subseteq \Omega, m_{\oplus}(A) = (m_1 \oplus m_2)(A) = \frac{(m_1 \cap m_2)(A)}{1 - m_{\cap}(\phi)}$$

$$[\text{degré de conflit}] m_{\cap}(\phi) = (m_1 \cap m_2)(\phi) = \sum_{B \cap C = \phi} m_1(B)m_2(C)$$

La somme orthogonale est possible seulement si le conflit entre les deux bba n'est pas total, c'est-à-dire si au moins deux éléments focaux de  $m_1$  et  $m_2$  n'ont pas l'ensemble vide pour intersection.

### 3.2.7 Combinaison de variables aléatoires et de distributions de possibilité par encadrement de fonctions de répartition

Les p-boîtes, qui, d'une manière générale, caractérisent l'incertitude épistémique (manque de connaissance), sont une famille de probabilités définies par deux distributions de probabilité haute et basse afin de modéliser l'imprécision due au manque de données et de connaissances ; par exemple les valeurs fixes d'un modèle, la date d'un événement passé, etc.

Une p-boîte généralisée  $[\underline{F}, \overline{F}]$  [Destercke et al., 2008] [Destercke et Dubois, 2009] sur un ensemble fini  $\chi$  est une paire de fonctions comonotoniques  $(\underline{F}, \overline{F})$ , c'est-à-dire que pour toute paire d'éléments  $x, y \in \chi$ , on a  $\underline{F}(x) < \underline{F}(y) \Rightarrow \overline{F}(x) < \overline{F}(y)$ . Cette paire de fonctions vérifie:  $\underline{F}: \chi \rightarrow [0,1]$  and  $\overline{F}: \chi \rightarrow [0,1]$  tel que  $\underline{F}$  est toujours inférieure à  $\overline{F}$  (i.e.  $\underline{F} \leq \overline{F}$ ) et qu'il existe au moins un élément  $x$  dans  $\chi$  tel que  $\underline{F}(x) = \overline{F}(x) = 1$ . Une p-boîte généralisée  $[\underline{F}, \overline{F}]$  génère un ensemble particulier de probabilités  $P_{[\underline{F}, \overline{F}]}$  telles que  $P_{[\underline{F}, \overline{F}]} = \{P \in \mathbb{P}_{\chi} / \underline{F}(x_i) \leq P(\{x_1, \dots, x_i\}) \leq \overline{F}(x_i)\}$  avec  $\mathbb{P}_{\chi}$  l'ensemble des mesures de probabilités sur  $\chi$ . Si  $\chi$  est l'ensemble des réels ( $\chi = \mathbb{R}$ ) et que les ensembles des  $A_i$  sont de type  $(-\infty, x_i]$  avec  $x_i < x_j$  lorsque  $i \leq j$ , on retrouve la théorie classique des p-boîtes, qui a été étudiée par de nombreux auteurs [Ferson et al., 2003] [Troffaes et Destercke, 2011]. Les p-boîtes interprètent les distributions de possibilité comme étant des familles de fonctions de répartition encadrées par une fonction haute et une fonction basse. Enfin, les liens entre les

différentes théories présentées sont expliqués dans [Destercke et al., 2007] et [Troffaes et al., 2013].

### 3.3 Conclusion : les limites des approches classiques de modélisation et de combinaison d'informations imparfaites

Nous venons de constater qu'il existe de nombreuses approches de traitement de l'incertitude. La tendance générale est d'étendre les techniques actuelles de modélisation de l'incertitude. De cette manière, elles deviennent de plus en plus complexes à mettre en œuvre. Cependant, laquelle de ces techniques pourrait simuler la réalité de manière opérationnelle tout en étant rigoureuse ? Comment combiner des données issues de sources hétérogènes avec différentes sources d'incertitude ?

Les dépendances entre les différentes formes d'incertitudes et d'imprécisions doivent être modélisées rigoureusement, de même que les problèmes de fausses extensions des possibles doivent être évités, afin de ne pas compter deux fois l'imprécision sur une même valeur, lors des calculs.

La théorie des sous-ensembles flous puis la théorie des possibilités ont élargi la portée des périmètres des études scientifiques qui supposaient que l'incertitude est essentiellement aléatoire. Ces théories permettent ainsi de traiter des cas plus généraux d'incertitude de données ou modèles empiriques. Dans la suite de ce document, nous proposerons une extension de la théorie de Dempster-Shafer basée sur le produit cartésien d'événements et/ou d'hypothèses appartenant à différents ensembles sur des données dont les types d'incertitudes sont hétérogènes. Nous verrons que cette approche de la fusion de connaissances implique de modéliser les données probabilistes et possibilistes en termes de masses focales ; nous les interpréterons donc comme des mesures de crédibilité, plausibilité, probabilité pignistique [Shafer, 1976] [Destercke et al., 2007] [Shapiro, 2012]. Une fois ces transformations effectuées, nous verrons que la combinaison rigoureuse de telles connaissances implique des retours en arrière au cours des calculs lors de la recomposition avec d'autres quantités présentant d'autres types d'imperfections (c'est-à-dire des calculs de probabilités et de possibilités à partir de masses focales), qui nécessitent des calculs très complexes et non opérationnels pour le décideur.



En ce qui concerne les théories basées sur les p-boxes, les décisions sont définies sur des critères tels que des distributions de fonctions hautes et basses, ce qui entraîne en général une infinité de distributions possibles, avec une grande variabilité en sortie du système et une complexité certaine dans les calculs. Bien que ces théories prennent en compte différents types d'imperfections de manière satisfaisante, en particulier pour des variables continues, il manque une méthode permettant d'effectuer les calculs de manière à la fois rigoureuse et efficace.

Nous nous limiterons aux sous-ensembles flous de type 1, considérant que nos résultats pourront ultérieurement être facilement étendus aux « interval T2FS » [Liang et Mendel, 2000].

Notre but ne consiste pas à fournir une unique interprétation de la fusion de données dont les imperfections sont hétérogènes, issues de différentes théories de l'incertain, mais plutôt de constituer un cadre formel permettant de propager rigoureusement les différents types d'incertitudes qui pourraient être interdépendants, sur lesquels une interprétation particulière de fusion de données pourra être appliquée. Enfin, concernant l'aide à la décision, notre approche permettra de remplacer  $m$  calculs numériques sur  $n$  hypothèses éventuellement croisées avec d'autres par un seul calcul formel, tout en s'intégrant au méta-modèle systémique que nous proposons dans le prochain chapitre.

## Troisième partie : contributions

## 4 Chapitre 4 : un méta-modèle systémique pour l'optimisation durable des systèmes socio-environnementaux

En premier lieu, nous présentons un modèle général d'optimisation durable « orienté métier » qui servira de base à nos contributions. Le modèle général que nous proposons permet à l'expert d'un domaine de modéliser un système complexe pour inclure les répercussions des simulations effectuées dans le système avec les interdépendances et rétroactions associées ainsi que des évaluations de la durabilité en vue de son optimisation. Cette vision systémique permet de modéliser le système à l'aide des fonctions entre lesquelles transitent des flux. A un système concret on associe son système informatisé d'optimisation durable. Pour ce faire, nous avons identifié trois sous-systèmes (information, prédiction, évaluation) auquel nous ajouterons un quatrième qui contrôle les trois autres.

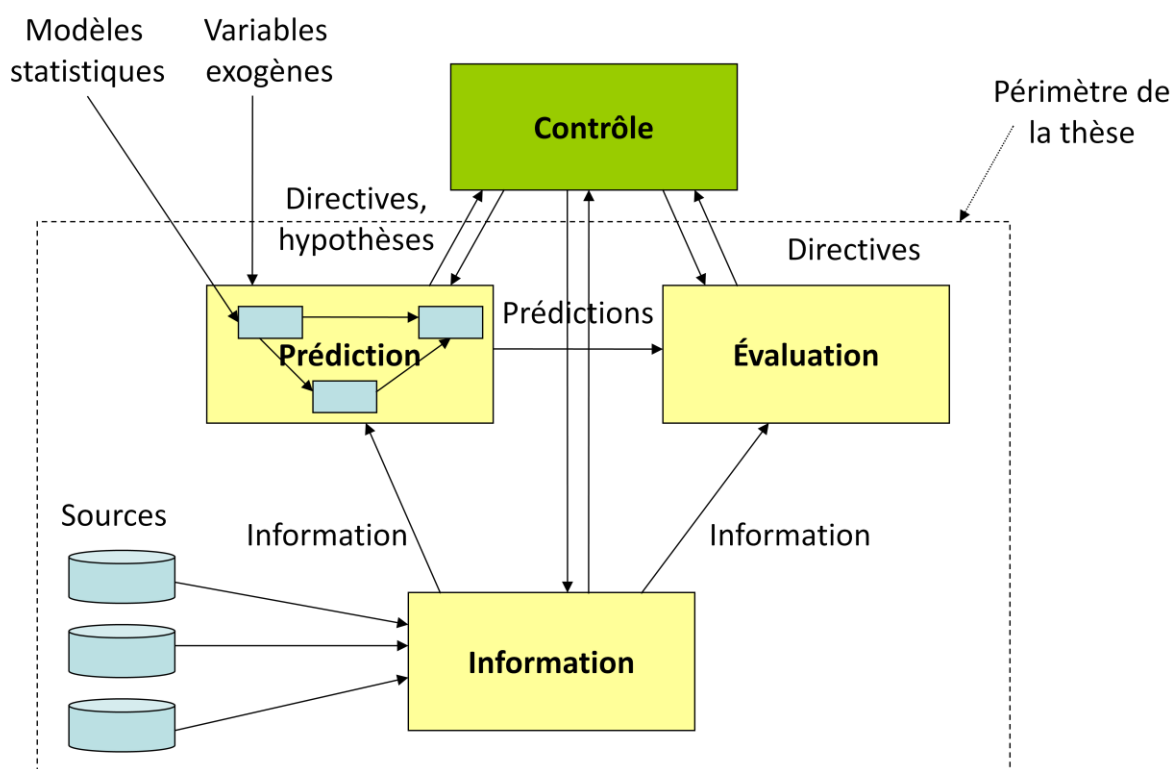


Figure 14 : le modèle général en « diamant »

Le premier sous-système concerne l'information, c'est-à-dire la modélisation et la collecte des informations nécessaires aux autres sous-systèmes, dotées de la précision requise.

Le deuxième sous-système se charge de la prédiction, c'est-à-dire de l'aspect simulation comportementale des fonctions opérationnelles du système. Le méta-modèle associé permet donc de modéliser le comportement des éléments du système et de leurs interactions au cours du temps, à l'aide de fonctions et de flux de données, au degré requis de précision. Le but n'est pas de fournir un modèle exhaustif d'un système complexe dans un domaine donné, mais de fournir un moyen aux experts du domaine de le modéliser, d'où la nécessité d'un méta-modèle.

Le troisième sous-système permet l'évaluation, c'est-à-dire de donner sens pour l'humain à des quantités physiques envoyées en sortie du système. C'est la traduction des données calculées qui se situent en sortie du système en indicateurs intelligibles par l'opérateur, afin que celui-ci prenne des décisions en connaissance de cause.

Un quatrième sous-système contrôle les trois autres. Il prend en charge l'optimisation du système global par coordination des trois autres. Par exemple, il peut tester des hypothèses, modifier des données d'entrée, reconfigurer des simulations, etc. Ce quatrième sous-système peut être un humain (c'est-à-dire un expert du domaine considéré). A terme, il s'agira d'un agent automatisé.

Dans la suite de ce document, nous ferons opérer ce modèle général « orienté métier » en synergie avec les concepts du domaine, les modèles fonctionnels et les éléments applicatifs du système socio-environnemental. C'est pourquoi, dans le prochain chapitre, nous définissons un méta-modèle.

#### 4.1.1 Un méta-modèle systémique dédié aux systèmes socio-environnementaux

Dans ce paragraphe, nous proposons un méta-modèle systémique des systèmes socio-environnementaux, qui est constitué de quatre modèles :

- Un ensemble de concepts génériques pour la modélisation du domaine de l'agriculture.
  - Un modèle structurel définissant les éléments structurels ainsi que les différents types d'imperfection des données rencontrées dans les systèmes socio-environnementaux.
- Le modèle d'imperfection des données circulant dans le système est indépendant des

types primitifs (entiers, réels). Il intègre des données pouvant être soit de nature continue soit de nature discrète.

- Un modèle fonctionnel dédié aux systèmes socio-environnementaux, en particulier aux exploitations agricoles. Ce modèle permet à l'expert de :
  - (1) Décrire les blocs fonctionnels présents (par exemple le sol, l'atmosphère, etc.) ou pouvant être présents (par exemple le bétail), avec les paramètres de configuration, dans une exploitation en général, entrées, sorties, etc., avec leurs relations possibles (par exemple, la production est dirigée vers un bloc « économie » qui convertit celle-ci en argent, et non l'inverse) ;
  - (2) Décrire la nature des flux (matière, énergie, informations) échangés entre les blocs et leur type ;
  - (3) Réaliser un modèle par l'expert du domaine : choisir les blocs génériques adéquats, les paramétrer, les agencer ;
  - (4) Fournir les implémentations concrètes (avec les domaines de validité des données, etc.).
- Un modèle des équations qui implémente les éléments du modèle fonctionnel.

Le méta-modèle que nous proposons est illustré dans la figure ci-dessous :

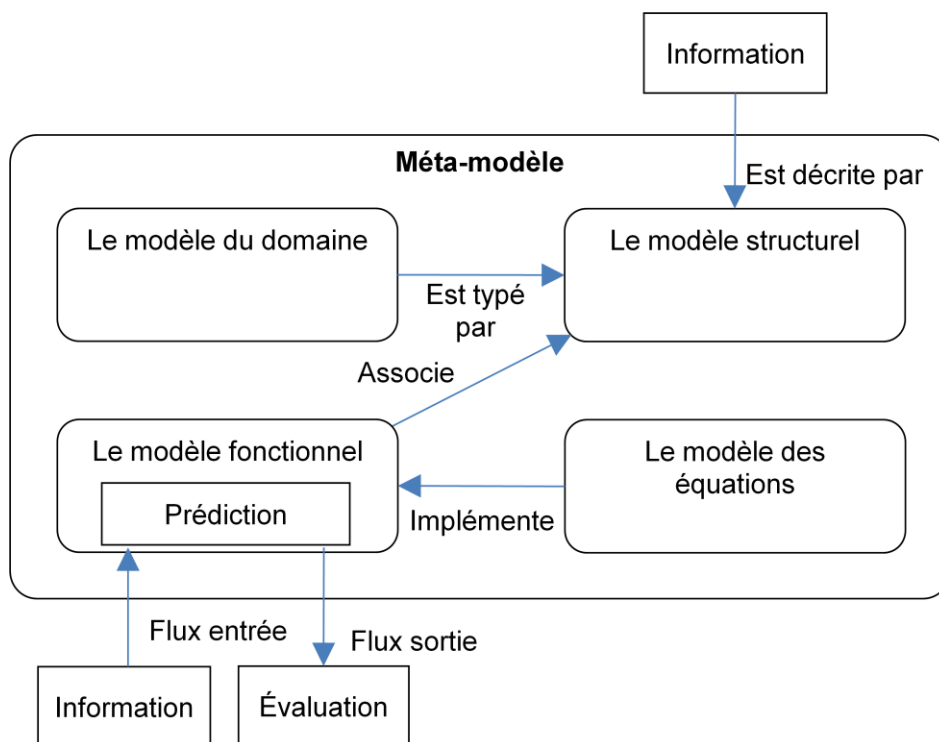


Figure 15 : le méta-modèle systémique - liens avec le modèle général

NB : la partie contrôle peut agir sur l'ensemble des modèles proposés (construction d'un modèle, paramétrage, etc.)

Nous montrerons que ce méta-modèle s'intègre dans le langage standard de modélisation des systèmes SysML sous forme d'extension (profil). Nous illustrerons cette extension par un exemple d'exploitation agricole intégrant le cycle de fertilisants circulant dans l'agrosystème.

#### 4.1.2 Modélisation du domaine : un modèle générique d'exploitation agricole

Nous définissons ci-dessous un modèle spécifique au domaine de l'agriculture. Nous modélisons une exploitation avec les éléments suivants :

(1) Des sources de données (paramètres) qui peuvent être internes ou externes à l'exploitation (le potentiel du sol, des indices de références scientifiques). Elles seront fournies au modèle via des interfaces. Les sources de données que nous avons identifiées sont :

- Les données internes, qui sont propres à l'exploitation, qui caractérisent et/ou circulent entre les entités réelles de l'exploitation, par exemple la matière organique et les nitrates échangés entre le sol et la plante, des mesures de tel ou tel polluant dans le sol, avec des imprécisions liées à l'appareil de mesure.
- Les données externes issues de phénomènes exogènes, par exemple des phénomènes physiques externes au système, et qui agissent directement sur l'exploitation. Elles sont en général issues de sources externes, et stockées en dehors de l'exploitation. Ces données sont souvent des variables aléatoires qui relèvent de distributions probabilistes, comme la météo, les cours de bourse, etc.
- Les données externes issues de références scientifiques. Ces données concernent des variables qui, en général, évoluent peu au cours d'un cycle. Elles peuvent être plus ou moins précises, e.g. les doses conseillées de tel ou tel fertilisant par les industriels, les besoins variables au cours du temps de certains types de cultures en azote, les surfaces plus ou moins précises de culture des différents types de semence, les coefficients de croissance de plante imprécis, etc.

(2) Des flux où transitent des objets caractérisés par des données qui peuvent être continues ou discrètes et qui peuvent être imparfaits : énergie, eau, fertilisant, pesticides.

(3) Des blocs fonctionnels qui transforment des flux d'entrée en flux de sortie :

- Sol (reçoit les intrants et en fournit une partie aux végétaux)

- Atmosphère (reçoit les éléments évaporés et fournit certains éléments au sol)
- Éléments produits (productions animale et/ou végétale)
- Parcelles (contiennent les éléments précédemment cités)
- Ingénierie agronomique (définit les besoins en éléments en fonction des données telles que les objectifs de production, les besoins des éléments productifs)
- Marché agricole (transforme la production en revenu financier)
- Machines (outils d'aide à la production)

(4) Des éléments dont les propriétés ont des imperfections représentées par des modèles de types de données abstraits. Ces éléments circulent dans les flux ou bien sont des propriétés de blocs. Exemple : surface d'une parcelle imprécise, modèles statistiques des récoltes des années antérieures.

(5) Des modèles analytiques issus de la recherche en agronomie. Ces modèles implémentent les blocs fonctionnels, et doivent prendre en compte les imperfections entachant des données, tout en propageant celles-ci rigoureusement via les opérations algébriques des modèles analytiques. Pour cela, nous proposerons des structures algébriques nécessaires dans le chapitre suivant sur l'imperfection des données. Enfin, les variables dépendent en général du temps, discrétisé avec un pas de temps égal à un cycle d'un an.

La figure ci-dessous illustre un modèle générique d'exploitation agricole :

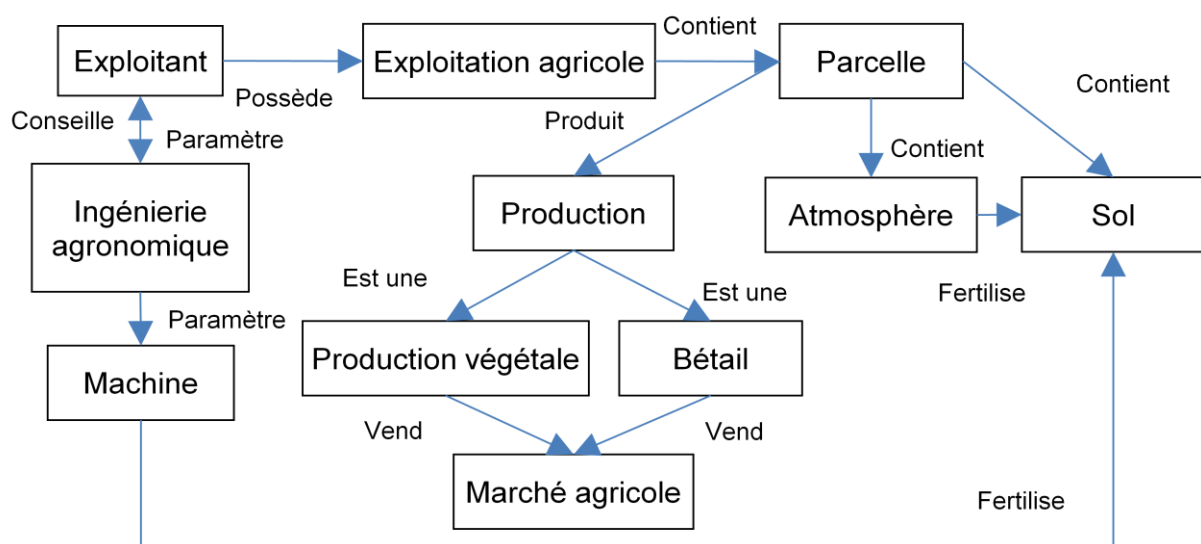


Figure 16 : le modèle générique d'exploitation agricole

#### 4.1.3 Un modèle structurel définissant les différents types d'imperfection des données rencontrés dans les systèmes socio-environnementaux

Dans ce paragraphe, nous proposons de modéliser les entités qui peuvent être soit statiques (propriété d'un élément ; par exemple le taux de pollution d'un sol) soit dynamiques (éléments et informations qui circulent dans le système).

Nous définissons des types d'attributs génériques exprimant différentes formes d'imperfections entachant les données ainsi que leur agrégation, caractérisant ainsi de manière générique les informations circulant au sein des systèmes socio-environnementaux. Nous avons vu que les données qui circulent dans le système sont imparfaites (imprécises, incertaines, plus ou moins probables), du fait de la source dont elles viennent (données exogènes), des calculs dont les résultats sont parfois des approximations, de la nature incertaine des prévisions futures (par exemple le temps qu'il fera), de leur nature d'hypothèse, etc. Il faut donc capturer des imperfections pour les propager correctement dans les calculs. Grâce au modèle de types de données abstraits que nous proposons, nous formaliserons les imperfections des données du modèle de l'agriculture que nous avons défini.

Nous nous limitons à quatre types de distributions afin de modéliser les types d'incertitude généralement rencontrés dans le cadre des systèmes socio-environnementaux. Nous avons choisi ces types de distributions non seulement pour leur expressivité (elles permettent en effet de modéliser un grand nombre de situations représentatives de la réalité) mais aussi pour leur implémentabilité (en effet, elles sont stables par un certain nombre d'opérations algébriques que nous présenterons dans le chapitre suivant dans lequel nous présentons un formalisme unifié pour modéliser l'imperfection des données). Les quatre types de distributions sont les suivants :

(1) « fuzzyDiscrete » : distribution discrète de valeurs floues. Une liste normalisée d'éléments flous discrets est définie par ses éléments et leurs degrés d'appartenance correspondants. La combinaison de tels éléments se fait par t-normes et t-conormes (e.g. min et max). On peut modéliser des avis d'experts vagues et/ou qualitatifs (linguistic predicates) ; on peut par exemple modéliser des expressions du type : (1) sans mesure préalable de la composition du sol permettant de se référer au triangle de texture de GEPPA [Baize, 1995], il est plutôt limoneux (possibilité 0.9), un peu sableux (possibilité 0.3) et n'est presque pas argileux (possibilité 0.1) ; (2) il est tout à fait possible que la récolte soit moyenne (possibilité



1), ou bien elle sera peut-être mauvaise (possibilité 0.2) mais certainement pas excellente (possibilité 0).

(2) « trapezoid » : un intervalle flou de type LR symétrique (i.e. un « trapézoïde » symétrique avec  $L = R$ ) est défini par sa valeur centrale, son noyau (valeur qui se situe avec certitude dans cet intervalle ; degré d'appartenance égal à 1) et son support (degré d'appartenance entre 0 et 1). Cette forme d'intervalle flou est très utilisée car de telles distributions sont quasi-stables par les opérations algébriques de base (addition, soustraction, multiplication, division). De plus, trois cas particuliers de cette forme, eux aussi très couramment utilisés, peuvent être mis en évidence : (1) des distributions de possibilité sous forme de rectangle (qui correspond à un noyau et un support identiques), (2) des distributions de possibilité sous forme de triangle (qui correspond à un noyau de largeur nulle), et (3) une valeur exacte et certaine égale à une valeur centrale avec un noyau et un support nuls. Grâce à ce type de distribution, on peut modéliser des valeurs floues de surfaces, de mesures scientifiques de la composition du sol, de doses d'azote de nourriture imprécises, etc.

(3) « probableDiscrete » : distribution de probabilité discrète. Les méthodes statistiques (échantillonnage, sondages, tests) sont des moyens puissants pour évaluer des probabilités d'événements.

(4) « normal » : une distribution de probabilité suivant la loi normale (gaussienne), définie par sa valeur centrale (qui correspond à la fois à son mode, sa moyenne et sa médiane) et sa variance. Par exemple, la masse des bovins d'un troupeau peut suivre une loi normale, de même que les récoltes au fil des ans. Ce type de distribution modélise en général des valeurs continues basées sur la répétition d'événements.

Nous définissons les quatre types de distributions décrits ci-dessus comme des « value types » abstraits (notés ADT<sup>26</sup> dans la suite de ce document). Ces ADT sont caractérisés par cinq propriétés : le nom, les types des données utilisés, les primitives, les préconditions et les axiomes. Le tableau suivant présente les éléments caractéristiques de ces types abstraits.

NB : ces ADT sont orthogonaux aux types primitifs (int, float, etc.). Ces derniers sont représentés par le terme « element ».

---

<sup>26</sup> Abstract Data Types

Tableau 8 : l'ADT « fuzzyDiscrete »

<b>Types utilisés</b>	List, element, float, element
<b>Primitives</b>	<pre>// Constructeur create: element, float → fuzzyDiscrete // Retourne les éléments qui ont un certain degré d'appartenance getElement: float → list[element] // Retourne tous les éléments de la distribution getElements: → list[element] // Ajoute un élément avec le degré d'appartenance correspondant addElement: element, float ∈ [0,1] → null // Retourne le degré d'appartenance d'un élément getGradeOfMembership: element → float ∈ [0,1] // Modifie le degré d'appartenance d'un élément setGradeOfMembership: element, float ∈ [0,1] → null // Supprime un élément deleteElement: element → null</pre>
<b>Pré conditions</b>	<pre>create: float=1 getElement: float ∈ [0,1] addElement: float ∈ [0,1] setGradeOfMembership: float ∈ [0,1]</pre>
<b>Axiomes</b>	

Tableau 9 : l'ADT trapezoid

<b>Types utilisés</b>	Float
<b>Primitives</b>	<pre>// Constructeur create : float, float, float, float → trapezoid // Retournent les valeurs définissant la distribution getCentralValue: → float getKernelMin: → float getKernelMax: → float getSupportMin: → float getSupportMax: → float // Modifient les valeurs définissant la distribution setKernelMin: float → float setKernelMax: float → float setSupportMin: float → float setSupportMax: float → float // Retourne les valeurs définissant la distribution getGradeOfMembership : float → float // Modifie la distribution de possibilité (ajout de connaissance)</pre>

	Modify : predicate $\rightarrow$ fuzzyFloat (intersection de deux)
<b>Pré conditions</b>	create : float, float, float, float: float < float < float < float setKernelMin: float < getKernelMax setSupportMin: float < getSupportMax setSupportMin: float < getKernelMin setKernelMax: float < getSupportMax
<b>Axiomes</b>	getKernelMin < getKernelMax getSupportMin < getSupportMax getSupportMin < getKernelMin getKernelMax < getSupportMax $\exists$ float $\in$ [getKernelMin, getKernelMax] / getGradeOfMembership = 1

NB : on peut facilement se ramener à des cas particuliers tels que la forme triangulaire (cas où getKernelMin= getKernelMax) et la forme rectangulaire (cas où getKernelMin= getSupportMin et getKernelMax= getSupportMax)

Tableau 10 : l'ADT probableDiscrete

<b>Types utilisés</b>	List, element, float, element
<b>Primitives</b>	// Constructeur create: element, float $\rightarrow$ probableDiscrete // Retourne les éléments qui ont une certaine probabilité getElement: float $\in$ [0,1] $\rightarrow$ list[element] // Retourne tous les éléments de la distribution getElements: $\rightarrow$ list[element] // Ajoute un élément avec la probabilité correspondante (attention, les probabilités des autres éléments sont repondérées) addElement: element, float $\rightarrow$ null // Retourne la probabilité d'un élément getProbability: element $\rightarrow$ float // Modifie la probabilité d'un élément setProbability: element, float $\rightarrow$ null // Supprime un élément (attention, les probabilités des autres éléments sont repondérées) deleteElement : element $\rightarrow$ null
<b>Pré conditions</b>	create: float=1 addElement: float $\in$ [0,1] setProbability: element $\in$ getElements, float $\in$ [0,1]
<b>Axiomes</b>	SUM(getElements.getProbability) = 1

Tableau 11 : l'ADT normal

<b>Types utilisés</b>	Float
<b>Primitives</b>	<pre>// Constructeur create: float, float→ normal // Retournent les valeurs définissant la distribution getAverage: → float getStandard_deviation: → float // Modifient les valeurs définissant la distribution setAverage: float → null setStandardDeviation: float → null //Calcule la probabilité d'un prédicat (en général un intervalle - calcul via la fonction de répartition) getProbability: predicate → float</pre>
<b>Pré conditions</b>	
<b>Axiomes</b>	

La figure suivante illustre le modèle des entités qui caractérisent les éléments statiques et qui circulent dans le système :

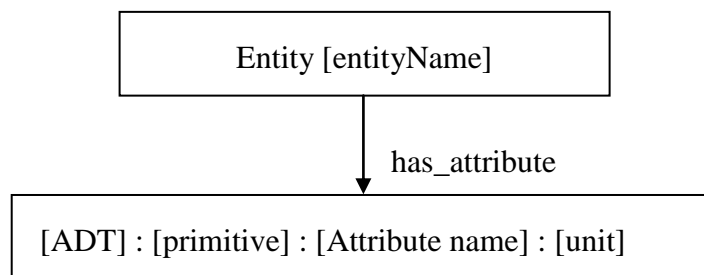


Figure 17 : le modèle des entités

#### 4.1.4 Le modèle fonctionnel

Dans ce paragraphe, nous allons présenter le modèle fonctionnel pour la prédiction des états futurs du système. Ce modèle définit les éléments simulés (exploitation, sol, atmosphère, etc.) avec les entrées et sorties typées par des flux de données discrètes (par exemple une tête de bétail) ou continues (par exemple de l'eau). Chaque bloc contient des attributs qui sont paramétrés le cas échéant par un ADT. L'ensemble des blocs fonctionnels reflète les éléments qui ont des relations systémiques entre eux. Les variables en entrée et en sortie des éléments systémiques sont interdépendantes. Les blocs en question comportent donc des inputs, des outputs et des connecteurs. Les fonctions réelles étant en général complexes, l'expert utilisera

des modèles analytiques tels que des linéarisations par des approximations tangentielles locales à des domaines de définitions précis.

Nous définissons un modèle fonctionnel (noté MF) comme un modèle qui décrit les actions du système étudié. Il est composé de fonctions, de flux entre ces fonctions, de blocs opérationnels, et enfin de relations d'opérationnalisation entre blocs et fonctions.

Nous définissons une fonction comme étant tout élément ou activité qui transforme des éléments d'entrée en éléments de sortie comme par exemple le sol, l'atmosphère, etc. En pratique, certaines fonctions peuvent appeler des sous-fonctions. Ainsi, la fonction de production d'une culture peut appeler une fonction de calcul de la fertilisation azotée. Une fonction  $F_i$  peut donc contenir des sous-fonctions  $F_{i,j}$ . Les sources de  $F_i$  et de  $F_{i,j}$  sont les entrées/sorties de la fonction parente  $F_i$ .

Entre les fonctions circulent des flux qui sont typés par des entités. Ces flux peuvent être continus (par exemple matière, eau, énergie, etc.) ou discrets (par exemple les ventes, la production, des informations, etc.). Les données circulant dans les flux sont imparfaites. Elles peuvent en effet être entachées d'incertitudes telles que l'imprécision, l'incomplétude ou avoir un caractère aléatoire. Ces incertitudes dépendent de leur source (dans le cas de données exogènes), des calculs numériques qui sont des approximations, de leur nature de prévision future (par exemple la météo prévue), de leur nature d'hypothèse, etc. La représentation mathématique ainsi que la propagation rigoureuse des imperfections dans les calculs seront abordées dans le prochain chapitre (« un formalisme unifié pour représenter l'information imparfaite »).

Des blocs opérationnels opérationnalisent les fonctions précédemment citées : un bloc est associé à une fonction avec (1) un domaine de validité pour les entrées (2) ainsi qu'une précision (par exemple un écart-type de la forme  $\frac{1}{|D|} \int_{Domaine} \varepsilon^2 \cdot d\varepsilon$ ). D'une manière générale, on dira qu'un bloc est une approximation locale d'une fonction, comme par exemple des équations différentielles linéarisées ou des régressions locales sur un domaine restrictif (par exemple une surface de l'exploitation agricole comprise entre 10 et 20 ha, etc.). Cette modélisation comportementale comporte des connecteurs permettant des rétroactions entre les différentes causes et effets des variables en jeu. L'agent superviseur peut ainsi agir ou fixer des paramètres afin d'observer le comportement du système sous certaines hypothèses. Les blocs sont paramétrés soit par des sorties d'autres blocs, soit par des variables externes. Nous

définirons ces paramètres et déterminerons leur nature imparfaite éventuelle, comme par exemple la pluviométrie qui est incertaine sur un intervalle.

Des relations d'opérationnalisation associent les blocs aux fonctions avec : (1) un domaine de validité  $I_{N_1} \times I_{N_2} \times \dots \times I_{N_n}$  et (2) une précision ou qualité (par exemple un écart-type de formule  $\frac{1}{|D|} \int_{Domaine} \varepsilon^2 \cdot d\varepsilon$ ).

Un MF est donc un 4-uplet  $(F, I, O, R)$ , où :

- $F$  est un ensemble  $\{F_i\}_{i \in \mathbb{N}}$  de fonctions (entrées, sorties, formule),
- $I$  est un ensemble de flux (nom,  $F_{depuis}$ ,  $F_{vers}$ , type),
- $O$  est un ensemble de blocs opérationnels (qui produisent un schéma de blocs fonctionnels),
- $R$  est un ensemble de relations  $\langle Bloc, Fonction, Domaine, Epsilon \rangle$ .

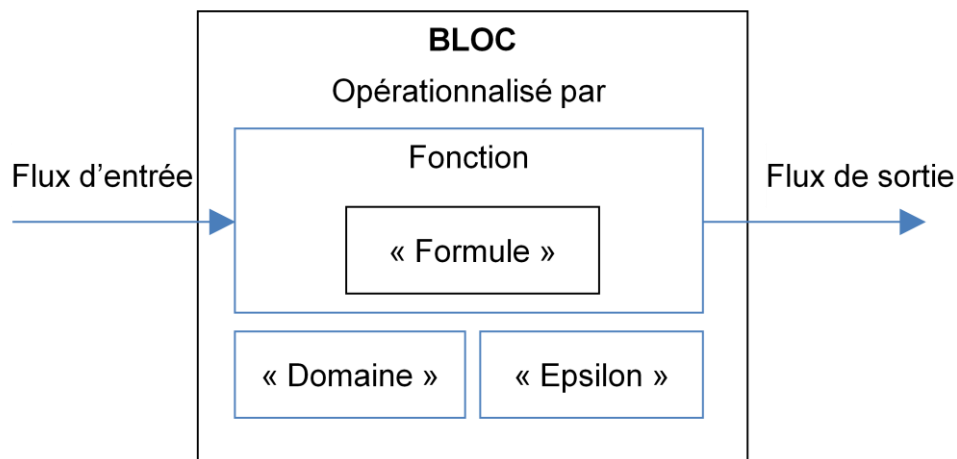


Figure 18 : schéma récapitulatif du modèle fonctionnel

En pratique les fonctions pourront contenir des sous-fonctions et appeler ces sous-fonctions. Dans ce cas, les sources et puits de la sous-fonction sont les entrées et sorties de la fonction parente.

#### 4.1.5 Intégration des nouveaux concepts au langage de modélisation standard des systèmes SysML

Nous proposons une correspondance naturelle entre les concepts du méta-modèle proposé et les notations classiques du langage standard de modélisation des systèmes SysML sous forme d'extension (profil SysML), tout en apportant la rigueur sémantique indispensable. Nous

illustrerons cette extension par un exemple d'exploitation agricole intégrant le cycle de fertilisants circulant dans l'agrosystème [Schvartz et al., 2005].

#### 4.1.5.1 Exemple de modélisation du cycle d'un fertilisant dans l'agrosystème

Dans cette étude, nous décrivons plus particulièrement le cas d'une exploitation agricole de grandes cultures mettant en œuvre le cycle d'un fertilisant ainsi que la production végétale et le revenu qui en résultent. Dans ce cas, les éléments qui circulent dans le système sont des objets « fertilisant » avec des champs « type\_fertilisant » et « quantité ». Les équations prennent un compte un pas de temps discret avec un pas de temps ( $\Delta t$ ) égal à un an. Le tableau suivant décrit les éléments du modèle en jeu :

Tableau 12 : modèle du domaine et fonctions spécifiques à l'agriculture

Fonction – Interne / externe	Obligatoire	Flux en entrée	Flux en sortie	Formule analytique en fonction d'un pas de temps discret (1an) – Domaine de validité
EXPL_AGR I (exploitation agricole) – Interne	Oui			
PARCEL – Parcelle	Oui	FertilisationIn (From : ITK, kg de fertilisant) (identique à MACHINE)	Production (To : ECO, kg végétal) (identique à la PV- production végétale)	
SOL – Interne	Oui	FertilisationIn (From : ITK, kg de fertilisant), Eau (From : ATM, kg de fertilisant)	FertilisationOut(To : CULTURE, kg de fertilisant), Volatilisation (To : ATM, kg de fertilisant)	Evaporation[N] + FertilisationOut[N] = FertilisationIn[N] + eau[N] – Reliquat[N-1] – $(\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+)$
ATM (Atmosphère) – Interne et externe	Oui	Evaporation (From : SOL, kg de fertilisant)	FertilisantEau (précipitations) (To : SOL, kg de fertilisant)	Evaporation[N] = FertilisantEau[N] (par hypothèse) – $\mathbb{R}^+$
ITK (ingénierie agronomique) – Interne et externe	Oui	Données objectif_production (From : PERSONNE, kg/ha de production végétale), Données b (besoin en fertilisant) (From : ITK, kg de fertilisant/kg de production végétale), Données surface (From : PARCELLE, ha) Données Revenu (From : EXPL, €) Données prix_marché (From : ECO, €/kg de production végétale)	Données fertilisation (To : MACHINE, kg de fertilisant)	Fertilisation[N] = b * Objectif_production[N] * surface – $0 < \text{Fertilisation} < 700$

Fonction – Interne / externe	Obl igat oire	Flux en entrée	Flux en sortie	Formule analytique en fonction d'un pas de temps discret (1an) – Domaine de validité
ECO (marché des matières premières agricoles) – Externe	Oui	Production (From : CULTURE, kg de production végétale)	Revenu (To : EXPL_AGRI, €)	Revenu[N] = Production[N] * prix_marché[N] – $(\mathbb{R}^+ \times \mathbb{R}^+)$
PV (production végétale) – Interne	Non	Fertilisation (From : SOL, kg de fertilisant), Eau (From : ATM, kg de fertilisant) prodParam (From : ITK, kg/ha)	Production (To : ECO, kg de production végétale)	Production[N] = X[N].( prodParam- X[N]) (avec X[N] = quantité de fertilisant et prodParam = paramètre propre au fertilisant et à la culture venant de ITK) – $0 < X < 800$
MACHINE (Machine à épandre) – Interne	Non	FertilisationIn (From : ITK, kg de fertilisant)	FertilisationOut (To : SOL, kg de fertilisant)	FertilisationOut[N] = FertilisationIn[N] – $\mathbb{R}^+$

#### 4.1.5.2 Lien entre le modèle structurel proposé et le diagramme de définition de bloc du langage standard de modélisation SysML

Les entités du modèle structurel statiques ou dynamiques s'associent respectivement de manière naturelle aux éléments suivants de SysML :

- Les blocs constituent les éléments de base pour la modélisation de la structure d'un système et sont caractérisés par des valeurs (« value properties ») qui sont utilisées pour modéliser les caractéristiques quantitatives des blocs en terme de types de valeurs (« value types »).
- Les flux ou « flow ports » autorisent la circulation de flux physiques entre les blocs. Les éléments échangés sont des objets modélisés par des blocs caractérisés par des attributs.

Nous proposons d'intégrer les ADT [Dantan et al., 2015 (1)] dans les diagrammes SysML suivants : le diagramme de définition de bloc, le diagramme de bloc interne et enfin le diagramme paramétrique. Le concepteur pourra donc spécifier des variables continues ou discrètes qui sont décrites par des données entachées des types d'imperfection suivants : imprécises, incomplètes et aléatoires. Le système simulé effectue des calculs (opérations algébriques classiques et opérateurs classiques d'agrégation) sur de telles quantités, propageant de manière rigoureuse les imperfections. Dans le chapitre suivant, nous présenterons les fondements mathématiques des opérations entre ADT.



Comme nous l'avons vu dans l'état de l'art, dans les diagrammes de blocs, les valeurs (« value properties ») sont utilisées pour modéliser les caractéristiques quantitatives des blocs, en terme de types de valeurs (« value types ») et basés sur les types de base proposés par SysML, notamment les distributions de probabilité (loi uniforme *<<uniform>>* et gaussienne *<<normal>>*) ainsi que les intervalles (*<<interval>>*). Nous avons présenté l'exemple de la surface du bloc « parcelle », typée par un « value type » « surface », défini par une distribution de probabilité à valeurs réelles, une dimension « surface » et une unité « hectare » ; soit :

```
<<uniform>>{min=10,max=12} surface:ha
```

Cependant, cette donnée de surface peut ne pas reposer sur une donnée statistique mais plutôt sur une incertitude subjective, qui de fait, se modélise de manière plus réaliste et rigoureuse par une distribution de possibilité (voir l'état de l'art).

La sémantique générale est définie comme suit :

```
<<distribution_name>>{parameters} {property name}: {unit name}
```

Les primitives des quatre types de valeurs distribuées, à savoir les distributions floues discrètes (*<<fuzzyDiscrete>>*) et continues (*<<trapezoid>>*) ainsi que lois de probabilité discrète *<<probableDiscrete>>* et normale *<<normal>>* dont nous avons besoin dans notre modèle agronomique, sont présentées ci-après :

```
<<fuzzyDiscrete>>{p1=(value1, gradeOfMembership1), p2=(value2, gradeOfMembership2),...} {property name}: {unit name}
```

```
<<trapezoid>>{KernelMin= value1, KernelMax= value2, supportMin= value3, supportMax= value4} {property name}: {unit name}
```

```
<<probableDiscrete>> {p1=(value1, probability1), p2=(value2, probability2),...} {property name}: {unit name}
```

```
<<normal>>{mode=m, standardDeviation=sd} {property name}: {unit name}
```

### *Exemples*

#### **L'ADT *<<fuzzyDiscrete>>* :**

(1) la production peut s'exprimer de la manière suivante : « il est tout à fait possible que la récolte soit moyenne (possibilité 1), ou bien elle sera peut-être mauvaise (possibilité 0.2) mais certainement pas excellente (possibilité 0) » :

```
<<fuzzyDiscrete>>{ (value="average", gradeOfMembership=1), ( value="bad", gradeOfMembership=0.2), (value="excellent", gradeOfMembership=0)}
production: kg
```

### **L'ADT <<trapezoid>> :**

(2) Nombre réel flou. La surface de la parcelle peut être quantifiée non pas avec une distribution de probabilité, mais avec l'expression suivante :

```
trapezoid{KernelMin=10, KernelMax=12, supportMin=10, supportMax=12}  
surface: ha
```

(3) Autre exemple (trapèze) :

```
trapezoid{KernelMin=9.5, KernelMax=12.5, supportMin=9, supportMax=13}  
surface: ha
```

(4) Ignorance totale. « On ne connaît pas du tout la surface de cette parcelle ». Autrement dit, toutes les valeurs sont possibles. Ceci peut être exprimé par la formule suivante :

```
trapezoid{mode=null} surface: ha
```

(5) Ajout de connaissances. On sait maintenant que « Surface2 » précédemment définie est supérieure ou égale à 12 ha.

```
trapezoid{KernelMin=9.5, KernelMax=12.5, supportMin=9, supportMax=13}  
surface: ha
```

La valeur actuelle est mise à jour par l'opérateur classique de logique floue « min » de l'intersection des deux distributions de possibilité décrites ci-dessous :

```
min(trapezoid{KernelMin=9.5, KernelMax=12.5, supportMin=9, supportMax=13},  
trapezoid{KernelMin=12, KernelMax=infinity, supportMin=12, supportMax=  
infinity}) surface: ha
```

C'est-à-dire, après calcul :

```
trapezoid{KernelMin=12, KernelMax=12.5, supportMin=12, supportMax=13},  
surface: ha
```

### **L'ADT <<probableDiscrete>> :**

La probabilité qu'il pleuve plus de 1000 mm dans l'année est de 0.3 ; la probabilité qu'il pleuve moins de 500 mm dans l'année est de 0.2 ; la probabilité qu'il pleuve plus de 500 mm et moins de 1000 mm dans l'année est de 0.5.

```
probableDiscrete{(blé, probability=0.7), (maïs, probability=0.3)} culture:  
null
```

### **L'ADT <<normal>> :**

Variable aléatoire réelle. La masse des bovins d'un troupeau est répartie selon une distribution normale de mode 700 et d'écart-type 200.

```
<<normal>>{average=700, standardDeviation=200} masseBovin: kg
```

Le diagramme SysML de définition de blocs suivant illustre la structure de l'exploitation agricole, adaptée au modèle du cycle d'un fertilisant. Ce diagramme intègre le formalisme tenant compte de l'imperfection des données que nous venons de présenter.

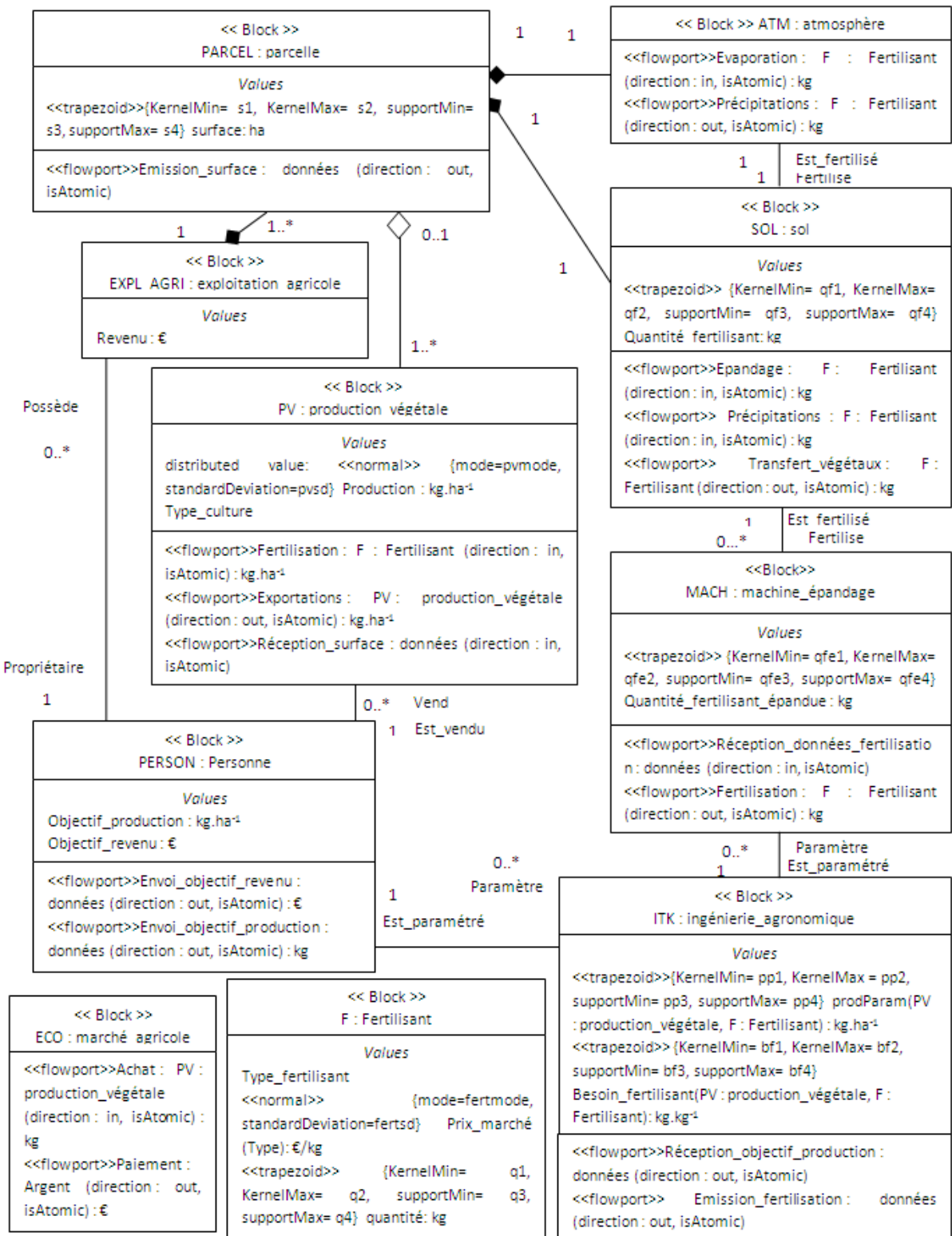


Figure 19 : diagramme de définition de bloc de l'exploitation agricole intégrant l'imperfection des données (cycle d'un fertilisant)

#### 4.1.5.3 Lien entre le modèle fonctionnel proposé et le diagramme de bloc interne du langage standard de modélisation SysML

Nous proposons une correspondance naturelle entre blocs et flux du modèle fonctionnel proposé et le diagramme de bloc interne du langage standard de modélisation SysML. Dans l'état de l'art sur la modélisation de systèmes via SysML, nous avons vu que les relations structurelles entre blocs telles que la composition (relation tout-partie) et l'agrégation (relation de contenance partagée) ainsi que l'association (référence) peuvent être associées à une cardinalité. Nous avons également vu que les interactions des ports du bloc englobant peuvent être déléguées aux ports des blocs fils via des délégations de ports. Les relations systémiques entre blocs fonctionnels via les flux nécessitent, dans certains cas, l'agrégation de plusieurs de ces flux (ceux-ci étant de même nature). En effet, dans le cas où les blocs fils n'ont pas la même cardinalité que le bloc père, SysML ne permet pas de formaliser cette différence de cardinalité lors de la prise en charge du port par le fils. Nous proposons de palier ce manque en intégrant la notion d'agrégation de flux (au sens comptable) via un mot clé « SUM » étiquetant le port.

Par exemple, le flux « PV : Production\_végétale » circule entre le bloc fonctionnel (de même nom) « PV : Production\_végétale » et le bloc « ECO : marché\_agricole » via plusieurs ports. Dans la réalité, le port situé entre le bloc « PV : Production\_végétale » et le port de sortie du flux vers le bloc « ECO : marché\_agricole » agrège les flux de productions végétales produites par le bloc « PARCEL : parcelle[\*] » à l'échelle de l'exploitation agricole entière. C'est pourquoi il est étiqueté avec le mot-clé « SUM ». On fait toujours la somme dans le sens « flux de l'élément contenant = somme des flux des éléments contenus ».

Le diagramme SysML de bloc interne suivant illustre les relations systémiques entre les blocs définis précédemment :

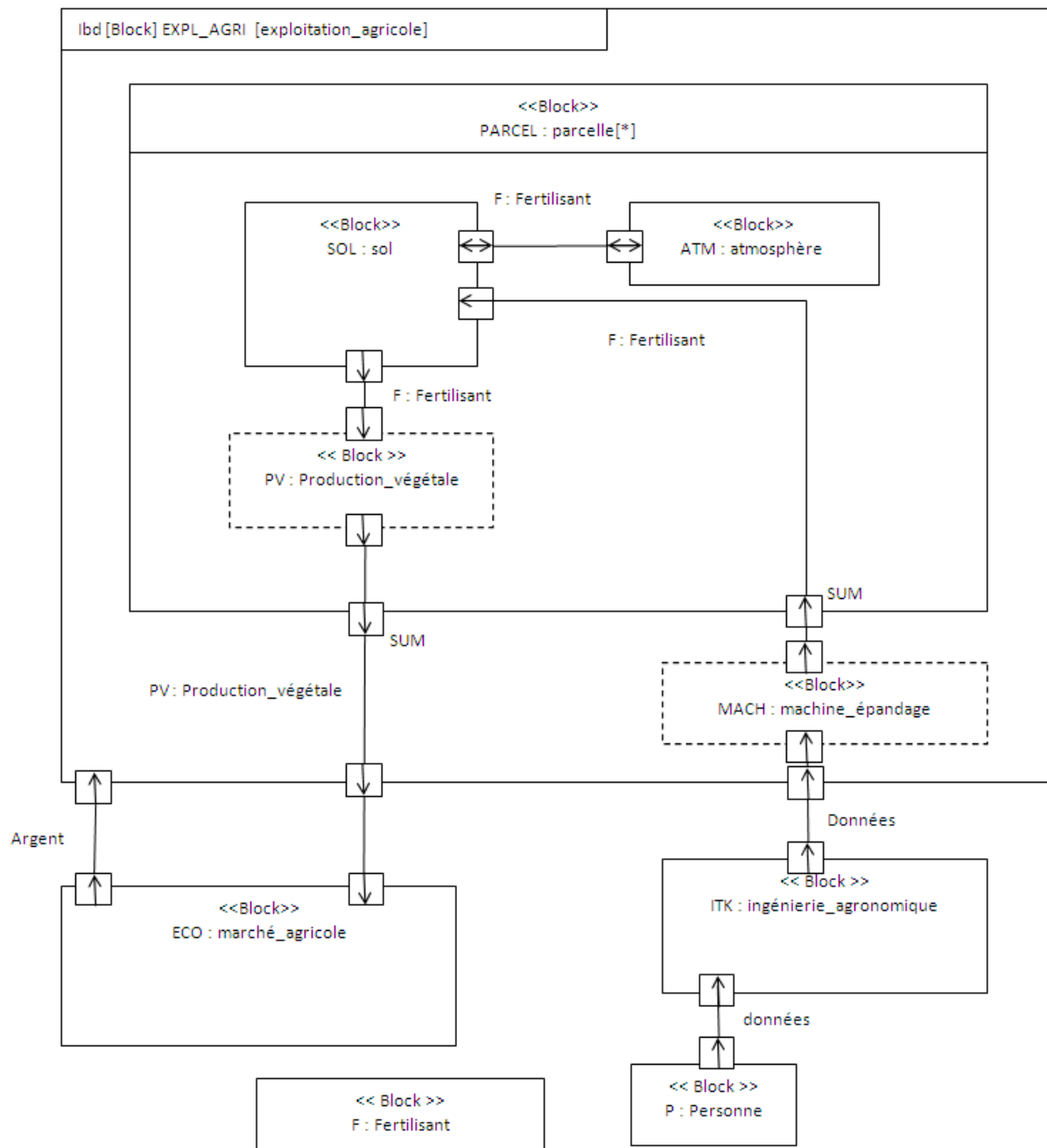


Figure 20 : diagramme de bloc interne de l'exploitation agricole intégrant l'imperfection des données (cycle d'un fertilisant)

#### 4.1.5.4 Implémentation du modèle fonctionnel via le diagramme paramétrique

Nous définissons un modèle d'équations qui instancie les implémentations des blocs. En effet, dans notre exemple, un tel modèle formalise le comportement du cycle d'un fertilisant dans l'exploitation via des équations. Dans ce cas, les blocs correspondent à des contraintes définies sous forme d'équations (on parle alors de « constraint block »). Dans le paragraphe

précédent, nous avons modélisé les différents types d'imperfections des paramètres et des variables en jeu dans ces équations (flous, probabilistes, ...). L'évaluation des équations mais aussi d'indicateurs nécessitent par conséquent de réaliser des calculs sur des données imparfaites, grâce à des opérateurs arithmétiques et/ou d'agrégation, comme une somme, un produit, une moyenne, etc. que nous devons conceptualiser puis implémenter de manière rigoureuse.

Nous définissons une implémentation d'un bloc fonctionnel, comme un diagramme paramétrique contenant un graphe d'équations liées entre elles par les liaisons appelées bindings.

Par exemple, la production totale se calcule à partir du produit d'une surface et d'une récolte qui peuvent être respectivement imprécise et aléatoire. De plus, sur une exploitation comportant plusieurs parcelles, elle se calcule à partir de la somme des productions des parcelles où sont cultivées ces mêmes cultures.

Autre exemple : l'équation qui détermine le comportement du peuplement végétal cultivé a pour entrée la quantité de fertilisant fournie par la fertilisation et la pluie et pour sortie la production végétale. Elle détient pour paramètre la surface cultivée, qui peut être imprécise.

Nous pouvons lier une variable avec les trois éléments suivants, sous réserve de cohérence des types de données :

- Un champ situé en entrée d'un bloc interne (IN),
- Un champ situé en sortie d'un bloc interne (OUT),
- Une propriété d'un bloc structurel.

Nous intégrons enfin la temporalité éventuelle pour chaque variable :  $[N]$  signifie cycle (annuel dans cet exemple) en cours.

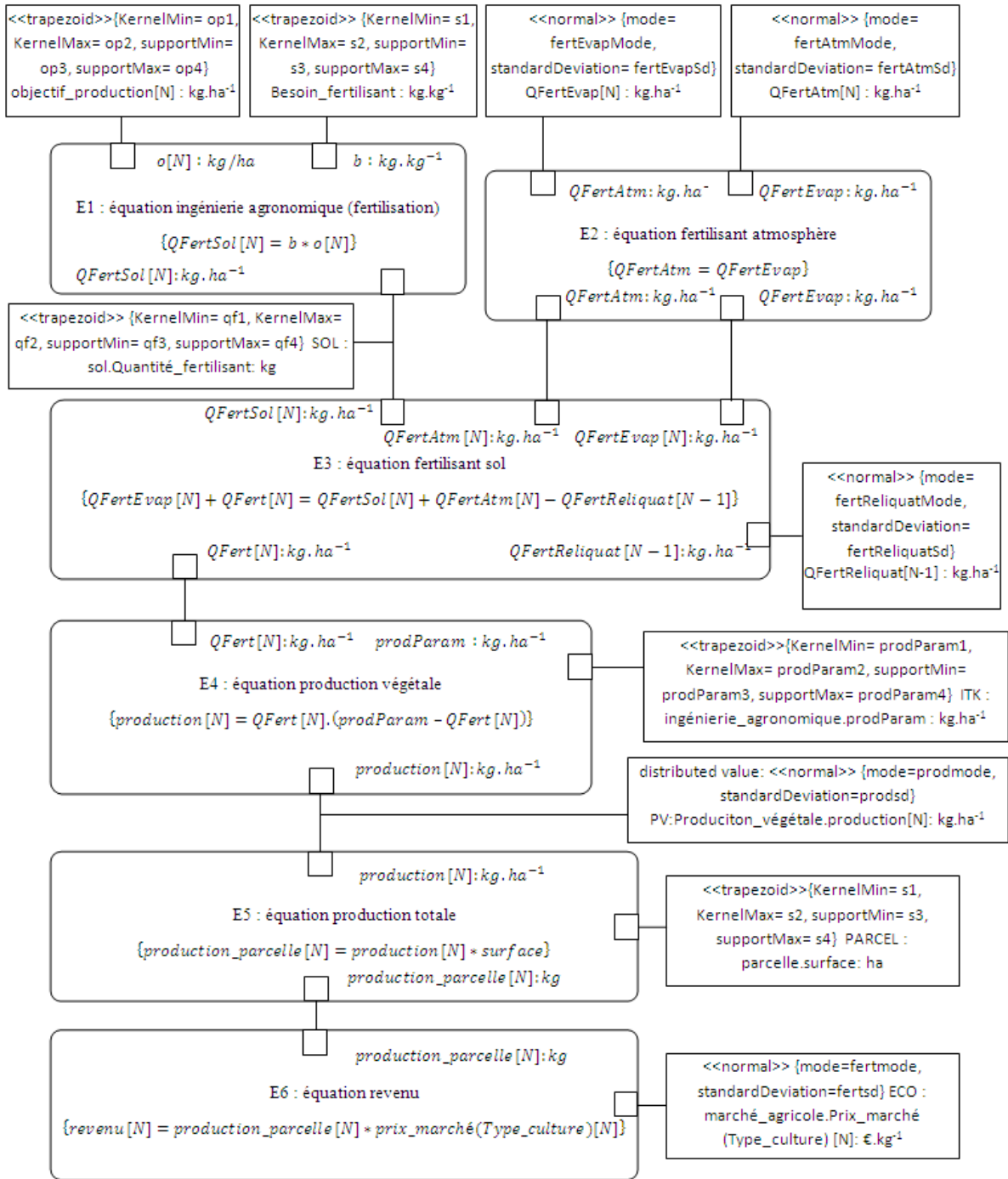


Figure 21 : diagramme paramétrique de l'exploitation agricole intégrant l'imperfection des données (cycle d'un fertilisant)

#### 4.1.6 Conclusion : un méta-modèle systémique nécessitant un formalisme unifié pour représenter et opérer sur l'information imparfaite

Dans ce chapitre, nous avons proposé un méta-modèle systémique pour la représentation de systèmes socio-environnementaux. Ce méta-modèle est constitué de quatre modèles : le modèle du domaine, le modèle structurel, le modèle fonctionnel et le modèle des équations. Ces éléments sont plus particulièrement :

- Le modèle du domaine : un ensemble de concepts pour la modélisation du domaine de l'agriculture.
- Le modèle structurel : ce modèle définit des entités ainsi que des types d'attributs génériques exprimant différents types d'imperfection.
- Le modèle fonctionnel : ce modèle définit des systèmes socio-environnementaux génériques via des flux et des blocs fonctionnels.
- Le modèle des équations : il s'agit de l'implémentation du modèle fonctionnel par le diagramme paramétrique SysML.

Nous avons intégré les nouveaux concepts du méta-modèle proposé au langage de modélisation standard SysML (profil).

Les opérateurs arithmétiques sur les types abstraits que nous avons définis, prenant en compte différents types d'imperfections sur les données, nécessitent un formalisme unifié pour représenter et opérer sur de telles informations. Nous allons proposer ces contributions dans le chapitre qui suit.



## 5 Chapitre 5 : un formalisme unifié pour représenter et opérer sur l'information imparfaite

### 5.1 Introduction

Dans ce chapitre, nous présentons notre seconde contribution : un formalisme unifié pour modéliser et manipuler les informations dans les différentes formes d'imperfection rencontrées dans le cadre des systèmes socio-environnementaux [Dantan et al., 2015 (2)] [Dantan et al., 2014 (2)].

Dans le chapitre précédent, nous avons vu que, d'une part les attributs de blocs et que, d'autre part, les flux d'objets sont paramétrés par des ADT qui se caractérisent par des quantités de nature continue ou discrète et peuvent se révéler imparfaits. Nous avons identifié quatre types d'imperfection : Fuzzydiscrete, FuzzyContinuous<sup>27</sup>, ProbableDiscrete, Normal. Ces types d'imperfections sont orthogonaux aux types de données qui peuvent être des valeurs de types primitifs (nombres entiers, nombres réels), des hypothèses ou même des collections d'éléments tels que des attributs multivalués ou des objets.

Le formalisme que nous proposons permet de gérer des quantités qui peuvent prendre un nombre fini de valeurs dont la valeur elle-même peut ne pas être connue précisément (imprécision, variabilité) et dont l'existence peut être incertaine (possible ou probable).

Dans ce chapitre, nous fournissons :

- Un modèle de représentation de telles données, préservant la dépendance entre sources d'incertitudes
- Un modèle de calculs rigoureux afin de réaliser des opérations algébriques sur de telles quantités.
- Une structure informatique pour le stockage des résultats des modèles précédemment définis.

---

<sup>27</sup> Ici par abus de langage ce qu'on appelle fuzzy peut également être assimilé à une distribution de possibilité

## 5.2 De l'utilité de distinguer probabilités et possibilités dans un contexte de prise de décision

Le calcul d'incertitudes grâce à la théorie des possibilités ne permettent pas d'être plus précis. En effet, nous avons vu précédemment que, dans certaines approches, une distribution de possibilité est interprétée comme une probabilité imprécise ou comme une famille de distributions de probabilité. Cependant, dans certains cas, les résultats modélisent l'imprécision de manières plus « logique » et rigoureuse. De plus, les calculs sont alors plus simples à effectuer.

En effet, considérons une série de mesures imprécises  $(a_i)_{i=1}^{i=n}$  réelles dont l'imprécision est constante si l'appareil est identique entre deux mesures. Cependant, si l'on change d'appareil entre deux mesures, il faudrait prendre en compte une valeur d'imprécision différente.

(1) Si nous utilisons la théorie des probabilités, nous choisissons classiquement de modéliser les valeurs  $a_i$  obtenues de chaque mesure et leur imprécision  $\delta a_i$  par des lois uniformes centrées sur  $a_i$  de support  $[a_i \pm \delta a]$ . Cela signifie que nous avons fait l'hypothèse d'équiprobabilité à l'intérieur de chaque « intervalle ». Nous obtenons une série de mesures dont la distribution est de la forme  $(a_i, U(\delta a_i))_{i=1}^{i=n}$  avec les notations utilisées précédemment pour les « variateurs ». Les inconvénients de cette approche « probabiliste » sont les suivants :

- Tout d'abord, la loi de probabilité de l'ensemble des mesures est multimodale et est très compliquée à modéliser à l'aide d'une densité, surtout si les intervalles se recouvrent.
- Le calcul de l'espérance est complexe : l'approche consistant à calculer l'espérance des espérances de chaque distribution uniforme ne prend pas en compte l'étalement de chaque distribution (par exemple si l'appareil a été changé au cours de mes mesures). Pour tenir compte de l'étalement il faudrait calculer des sommes de lois uniformes. Or la somme de plusieurs lois uniformes nécessite des calculs complexes de probabilités. Par exemple, la somme de deux distributions uniformes a pour densité une loi triangulaire qui ne respecte donc pas l'hypothèse d'équiprobabilité (non stabilité des lois probabilistes en général par addition – sauf pour les gaussiennes dans ce cas – ni par multiplication).

- Enfin, il n'est pas vraiment possible d'intégrer la notion de crédibilité d'une source de manière stable. Mettre une probabilité sur l'existence d'une mesure « intervalle » consisterait par exemple à modéliser une série de mesures par l'expression suivante :  $\sum_{i=1}^{i=n} P(a_i) \cdot a_i \cdot U[\delta a_i]$ . C'est-à-dire qu'il y a  $(1 - P(a_i))$  risque que  $a_i$  soit complètement fausse. Cette représentation est alors potentiellement instable (tout ou rien).

(2) La théorie des possibilités permet d'aboutir à des modèles plus simples à calculer, surtout lorsque le nombre de mesures n'est pas suffisant pour appliquer la loi des grands nombres :

- Classiquement nous modélisons les valeurs  $a_i$  obtenues de chaque mesure soit par des intervalles (logique classique) formant des « rectangles » soit par des sous-ensembles flous des valeurs possibles de la mesure (logique floue) formant des « trapézoïdes », intégrant ainsi de la nuance dans la représentation des possibilités. Nous obtenons une série de mesures dont la distribution est soit de la forme  $(a_i \cdot T(\delta a_i))_{i=1}^{i=n}$  (rectangles) soit  $(a_i \cdot T(\delta a_i, \delta a'_i))_{i=1}^{i=n}$  (trapézoïdes) avec les notations utilisées précédemment pour les « variateurs ».
- Les valeurs ainsi définies se combinent arithmétiquement par les opérateurs classiques de la logique floue (exemple : T-norme min et T-conorme max). Comme nous l'avons vu précédemment, les formes trapézoïdales prennent en compte l'étalement de chaque distribution et sont stables par les opérations arithmétiques classiques (même pour des trapèzes qui se recouvrent) : les calculs reviennent à effectuer des calculs sur des intervalles pour chaque alpha-coupe. L'hypothèse d'équipossibilité est donc conservée.
- La crédibilité de sources d'information  $(a_i)_{i=1}^{i=n}$  peut facilement être graduée par le flou. Exemple :  $(a_i / \mu(a_i))_{i=1}^{i=n}$ .

Grâce à la théorie des possibilités, nous obtenons ainsi un modèle plus fidèle à la réalité et plus simple à mettre en œuvre.

## 5.3 Formalisation et calculs génériques sur des connaissances probabilistes et possibilistes discrètes

### 5.3.1 Notations et définitions

Considérons  $\Omega$  comme étant l'univers des possibles, doté d'une mesure de possibilité  $\Pi$  et d'une mesure de probabilité  $P$ , ayant chacune un nombre fini de valeurs. Ces valeurs peuvent appartenir soit à un corps  $K$  (par exemple  $\mathbb{R}$ ) soit à un demi-groupe  $G$  (i.e. un ensemble muni d'une loi de composition interne associative). Ces quantités peuvent être par exemple des valeurs de types primitifs (nombres entiers, nombres réels), des hypothèses ou même des collections d'éléments tels que des attributs multivalués ou des objets.

#### *Bases possibilistes et probabilistes*

Soit  $E$  un espace vectoriel sur  $K$ , de dimension infinie mais dénombrable. Soient  $B_I$  et  $B^J$  deux ensembles de bases de  $E$  avec :

- $B_I = \{X_{I,i}\}_{i=1,\dots,m}$  ( $I \in \mathbb{N}$  fixé,  $m \in \mathbb{N}$ ), que nous appelons bases possibilistes.
- $B^J = \{X^{J,j}\}_{j=1,\dots,n}$  ( $J \in \mathbb{N}$  fixé,  $n \in \mathbb{N}$ ), que nous appelons bases probabilistes.

On définit un produit interne « . » sur les vecteurs définis ci-dessus. Ils vérifient les propriétés suivantes :

- $X.X = X$  (idempotence)
- $X.1_v = X$  (élément neutre)
- $X.0_v = 0_v$ , où  $0_v$  est le vecteur nul (élément absorbant)
- $i_1 \neq i_2 \Rightarrow X_{I,i_1}.X_{I,i_2} = 0_v$
- $j_1 \neq j_2 \Rightarrow X^{J,j_1}.X^{J,j_2} = 0_v$
- $X_{I_1,i_1}.X_{I_2,i_2} \neq 0_v$
- $X_{I_1,i_1}.X^{J_2,j_2} \neq 0_v$
- $X^{J_1,j_1}.X^{J_2,j_2} \neq 0_v$

Soit  $D_p(G)^{28}$  l'ensemble engendré par les vecteurs probabilistes définis ci-dessus.  $D_p(G)$  est donc l'ensemble fini des valeurs probabilistes sur  $G$  définies sur  $\Omega$ .

---

<sup>28</sup>  $D_p(G)$  : distribution probabiliste sur  $G$

Soit  $D_\pi(G)^{29}$  l'ensemble engendré par les vecteurs possibilistes définis ci-dessus.  $D_\pi(G)$  est donc l'ensemble fini des valeurs possibilistes sur  $G$  définies sur  $\Omega$ .

Soit  $D_I(G)^{30} = D_p(G) \cup D_\pi(G)$ . C'est l'ensemble engendré par les vecteurs probabilistes et possibilistes définis ci-dessus.  $D_I(G)$  est donc l'ensemble fini des valeurs probabilistes et possibilistes sur  $G$  définies sur  $\Omega$ .

### 5.3.2 Quantités possibilistes discrètes

#### 5.3.2.1 Loi de composition interne $+_\pi$

**Définition :** nous définissons une loi de composition interne sur  $D_\pi(G)$ , notée  $+_\pi$ , comme l'union de deux quantités possibilistes formalisée par l'expression suivante :

$$\forall (a_1, a_2) \in G^2, \alpha_{a_1 +_\pi a_2} = \max(\alpha_{a_1}, \alpha_{a_2})$$

**Définition :** nous définissons la forme canonique purement possibiliste de  $a \in D_\pi(G)$  par l'expression suivante :

$$a = a_1/\alpha_1 \cdot X_{I,1} +_\pi \dots +_\pi a_n/\alpha_n \cdot X_{I,n}$$

C'est-à dire, plus formellement :

$$a = \sum_{i=1}^n a_i/\alpha_i \cdot X_{I,i}$$

Avec :

- $a_i$  les valeurs possibles de  $a$ .
- $\alpha_i$  les possibilités associées à chaque valeur possible  $a_i$  (dont l'une au moins est égale à 1).
- Par hypothèse  $\forall i, j \in \mathbb{N}, a_i \cap a_j = \emptyset$  et  $a = \sum_{i=1}^n a_i/\alpha_i \cdot X_{I,i}$ . Les  $a_i/\alpha_i \cdot X_{I,i}$  (I fixé) représentent la partition de l'univers des possibles  $\Omega$  correspondant aux valeurs de  $a$ .
- **Contrainte de normalisation :**  $\max_{i=1}^{i=n}(\alpha_i) = 1$

---

<sup>29</sup>  $D_\pi(G)$  : distribution possibiliste sur  $G$

<sup>30</sup>  $D_I(G)$  : distribution incertaine sur  $G$

L'écriture de  $a$  est unique, les  $a_i$  (resp  $\alpha_i$ ) sont dits valeurs canoniques (respectivement poids canonique de  $a$ ).  $a$  s'interprète comme une valeur susceptible de prendre la valeur  $a_i$  avec la possibilité associée  $\alpha_i$ . Il s'agit ici typiquement d'une disjonction discrète et finie d'éléments de  $G$ , chacun étant associé à un degré de possibilité. Autrement dit, il s'agit de variables possibilistes sur  $\Omega$ , prenant un nombre fini de valeurs  $(a_i)_{i=1}^n \in G^n$ .

*Propriétés de  $(D_\pi(G), +_\pi)$*

- $0_v$  est l'élément neutre pour  $+\pi$ .
- $+\pi$  est commutative et associative.
- Il n'existe pas d'inverse pour  $+\pi$ .
- $(D_\pi(G), +_\pi)$  est un demi-groupe unitaire (ou monoïde) commutatif.

*Exemple*

Réel/entier flou. Soit  $a$  un nombre flou tel que : il est possible que  $a$  soit égal à 2, ou il est possible (0.8) que  $a$  soit égal à 3. La forme de connaissance s'exprime ainsi :

- $a = 2$  (possibilité 1) ou 3 (possibilité 0.8)

$a$  s'exprime dans la base canonique avec la formule suivante :

- $a = 2/1 \cdot X_{I,1} + 3/0.8 \cdot X_{I,2}$

On pourra également exprimer des combinaisons d'éléments plus ou moins possibles ou nécessaires :

$$a_1 = a_{1,1}/\alpha_{1,1} \cdot X_{I,1} + \emptyset/\alpha_{1,2} \cdot X_{I,2}$$

où  $\alpha_{1,1}$  exprime la possibilité et  $(1 - \alpha_{1,2})$  est la nécessité de la valeur  $a_{1,1}$ .

### 5.3.2.2 Loi de composition interne $\times_\pi$

**Définition :** soient  $(a_1, a_2) \in D_\pi(G)^2$  deux quantités purement possibilistes non interactives :  $a_1 = \sum_{i=1}^{n_1} a_{1,i}/\alpha_{1,i} \cdot X_{1,i}$  et  $a_2 = \sum_{j=1}^{n_2} a_{2,j}/\alpha_{2,j} \cdot X_{2,j}$ . Nous définissons une loi de composition interne sur  $D_\pi(G)$  notée  $\times_\pi$  formalisée par l'expression suivante :

$$\forall (a, a_1, a_{2,j}) \in G^3, \forall i, j \in \mathbb{N}^2, \alpha_{a_1 \times_\pi a_2}(a) = \sup_{a_{1,i} \times_\pi a_{2,j}} (\alpha_{1,i}(a_{1,i}) \min \alpha_{2,j}(a_{2,j}))$$

L'expression suivante :

$$a_1 \times a_2 = (a_{1,1}/\alpha_{1,1} \cdot X_{1,1} +_{\pi} \dots +_{\pi} a_{1,n_1}/\alpha_{1,n_1} \cdot X_{1,n_1}) \times_{\pi} (a_{2,1}/\alpha_{2,1} \cdot X_{2,1} +_{\pi} \dots +_{\pi} a_{2,n_2}/\alpha_{2,n_2} \cdot X_{2,n_2})$$

est évaluée avec une somme composée des  $(a_{1,i} \times_{\pi} a_{2,j})/(\alpha_{1,i} \min \alpha_{2,j}) \cdot X_{1,i} \cdot X_{2,j}$ .

*Démonstration*

$$\begin{aligned} a_1 \times_{\pi} a_2 &= \sum_{i=1}^{n_1} a_{1,i}/\alpha_{1,i} \cdot X_{1,i} \times \sum_{j=1}^{n_2} a_{2,j}/\alpha_{2,j} \cdot X_{2,j} \\ a_1 \times_{\pi} a_2 &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (a_{1,i}/\alpha_{1,i} \cdot X_{1,i} \times a_{2,j}/\alpha_{2,j} \cdot X_{2,j}) \end{aligned}$$

En appliquant la loi de composition interne  $\times_{\pi}$ , on obtient :

$$a_1 \times_{\pi} a_2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (a_{1,i} \times a_{2,j})/(\alpha_{1,i} \min \alpha_{2,j}) \cdot X_{1,i} \cdot X_{2,j}$$

Ce type de structure correspond au produit cartésien d'éléments de l'ensemble G (sous forme de disjonction de conjonction de valeurs de G), chacun étant affecté à une possibilité. Si  $a_1$  et  $a_2$  sont compatibles algébriquement, on pourra agréger leurs valeurs avec des lois de compositions internes telles que  $+$  ou  $*$ .  $D_{\pi}(G)$  héritera ainsi de certaines propriétés de G pour les lois  $+$  et  $*$ . On vérifie que l'on peut immerger l'ensemble G dans  $D_{\pi}(G)$ .

*Généralisation*

On déduit immédiatement la formule suivante pour n combinaisons d'éléments de  $G^n$  :

$$a_1 \times_{\pi} \dots \times_{\pi} a_n = \left( \prod_{k=1}^n a_i \right) = \sum_{i_1=1}^{N_1} \dots \sum_{i_n=1}^{N_n} \left( \prod_{k=1}^n a_{k,i_k} \right) / \min(\alpha_{1,i_1}, \dots, \alpha_{n,i_n}) \cdot \left( \prod_{p=1}^n X_{p,i_p} \right)$$

*Expression générale d'une distribution possibiliste sur G*

D'une manière générale, en considérant que chaque terme  $a_i$  est une agrégation d'éléments de G par la loi de composition interne  $\times_{\pi}$ , tout élément de  $D_{\pi}(G)$  est une expression de la forme :

$$a = \sum_{i=1}^n a_i / (\alpha_i \cdot \prod_{l=1}^{n_1} X_{l,i})$$

### *Cas particulier de deux opérandes exprimés dans la même base*

Si deux opérandes  $a_1$  et  $a_2$  sont exprimés sur la même base (c'est-à-dire qu'ils sont dépendants, ou en d'autres termes leurs dépendances sont liées), alors la formule se simplifie. En effet, intuitivement,  $a_1$  et  $a_2$  sont issus de la même source d'incertitude, i.e. lorsque le premier opérande vaut  $a_{1,i}$ , la seconde vaut  $a_{2,j}$  avec le même degré de possibilité.

On pose :  $\forall i \in \mathbb{N}, X_{1,i} = X_{2,i}$  et  $n_1 = n_2$  (bases identiques). On a donc :

$$\begin{aligned} a_1 \times_{\pi} a_2 &= \sum_{i=1}^{n_1} a_{1,i}/\alpha_{1,i} \cdot X_{1,i} \times_{\pi} \sum_{j=1}^{n_1} a_{2,j}/\alpha_{2,j} \cdot X_{1,j} \\ a_1 \times_{\pi} a_2 &= \sum_{i=1}^{n_1} (a_{1,i}/\alpha_{1,i} \cdot X_{1,i} \times_{\pi} a_{2,j}/\alpha_{2,j} \cdot X_{1,j}) \\ a_1 \times_{\pi} a_2 &= \sum_{i=1}^{n_1} (a_{1,i} \times_{\pi} a_{2,j}) / \text{Sup}(\alpha_{1,i}, \alpha_{2,j}) \cdot X_{1,i} \end{aligned}$$

Remarque : ce cas particulier prend en compte rigoureusement les dépendances entre sources d'incertitudes afin de ne pas augmenter de manière artificielle le nombre de valeurs possibles.

### *Propriétés de $(D_{\pi}(G), \times_{\pi})$*

- $\times_{\pi}$  est associative et distributive sur  $+$ .
- $(D_{\pi}(G), \times_{\pi})$  est un demi-groupe.

Considérons que  $\times_{\pi}$  peut être soit l'opération algébrique  $+$  (addition) soit l'opération algébrique  $*$  (produit). Alors  $D_{\pi}(G)$  hérite de certaines propriétés de  $G$ . Supposons que  $(G, +)$  et  $(G, *)$  sont des demi-groupes commutatifs unitaires, i.e.  $G$  possède deux éléments neutres pour  $+$  et  $*$  respectivement notés  $0_G$  et  $1_G$ .

- $+$  et  $*$  sont commutatives.
- L'élément neutre pour  $+$ , noté  $0_{D_{\pi}(G)}$ , est la variable possibiliste associant à tout élément de  $G$  la valeur  $0_G$ .
- Tout élément de  $D_{\pi}(G)$  possède un inverse par  $+$  :  $a = \sum_{i=1}^n a_i / (\alpha_i \cdot \prod_{l=1}^{n_1} X_{l,i})$  a pour inverse  $-a = \sum_{i=1}^n (-a_i) / (\alpha_i \cdot \prod_{l=1}^{n_1} X_{l,i})$ . Remarque : cette propriété ne serait pas valable si on considérait un élément de  $D_{\pi}(G)$  comme une simple distribution de



possibilité. Elle est valable parce qu'il s'agit d'une variable possibiliste tenant compte de la dépendance des incertitudes.

- $(D_\pi(G), \times_\pi)$  est un groupe commutatif unitaire.
- L'élément neutre pour  $*$ , noté  $1_{D_\pi(G)}$ , est la variable possibiliste associant à tout élément de  $G$  la valeur  $1_G$ .
- $0_{D_\pi(G)}$  est l'élément absorbant pour  $*$ .
- $*$  est distributive sur  $+$ . La propriété de distributivité de  $*$  sur  $+$  serait perdue si l'on considérait de simples distributions de possibilité sur  $G$  (et non des variables possibilistes). Dans notre cas, grâce à l'utilisation de vecteurs,  $(a + b) * a$  sera toujours égal à  $a * a + a * b$ .
- Tout élément de  $D_\pi(G)$  dont aucune valeur possible n'est nulle possède un inverse par  $*$  :  $a = \sum_{i=1}^n a_i / (\alpha_i \cdot \prod_{l=1}^{n_l} X_{l,i})$  a pour inverse  $1/a = \sum_{i=1}^n (1/a_i) / (\alpha_i \cdot \prod_{l=1}^{n_l} X_{l,i})$  avec  $(a_i)_{i=1}^n \neq 0_G$ . Un tel élément sera dit « nécessairement non nul ». Par contre aucun élément de  $D_\pi(G)$  « possiblement nul » (i.e. dont au moins une valeur possible est nulle) n'est inversible par  $*$ . Ainsi  $(D_\pi(G), +, *)$  ne possède pas une structure de corps.
- $(D_\pi(G), *)$  est un demi-groupe commutatif.
- $(D_\pi(G), +, *)$  possède une structure d'anneau commutatif unitaire.
- Si  $K$  est un corps,  $(D_\pi(G), +, \cdot)$  avec  $\cdot$  produit externe d'une distribution possibiliste d'éléments d'un corps  $K$  par un élément d'un corps  $K$  (loi de composition de  $K \times D_\pi(K) \rightarrow D_\pi(K)$ ) possède une structure d'espace vectoriel sur  $K$  ( $K$  peut par exemple être  $\mathbb{R}$ ).

### Exemples

(1) Soient  $a$  et  $b$  deux nombres flous. Il est possible que  $a$  soit égal à 2, ou il est possible (0.8) que  $a$  soit égal à 3. Il est possible que  $b$  soit égal à 1, ou il est possible (0.5) que  $b$  soit égal à 2.

La forme de connaissance s'exprime ainsi :

- $a = 2$  (possibilité 1) ou 3 (possibilité 0.8)
- $b = 1$  (possibilité 1) ou 2 (possibilité 0.5)

$a$  et  $b$  s'expriment dans la base canonique avec les formules suivantes, avec  $I \neq J$ :

- $a = 2/1. X_{I,1} + 3/0.8. X_{I,2}$
- $b = 1/1. X_{J,1} + 2/0.5. X_{J,2}$

En appliquant la loi de composition interne  $\times$ , on obtient :

$$a \times_{\pi} b = (2/1. X_{I,1} + \pi 3/0.8. X_{I,2}) \times_{\pi} (1/1. X_{J,1} + \pi 2/0.5. X_{J,2})$$

$$a \times_{\pi} b = (2 \times_{\pi} 1) / (1 \min 1). X_{I,1}. X_{J,1} + \pi (2 \times_{\pi} 2) / (1 \min 0.5). X_{I,1}. X_{J,2} + \pi (3 \times_{\pi} 1) (0.8 \min 1). X_{I,2}. X_{J,1} + \pi (2 \times_{\pi} 3) / (0.8 \min 0.5). X_{I,2}. X_{J,2}$$

(2) Dans le cas où  $\times_{\pi}$  est l'opérateur algébrique addition (« + »), on obtient :

$$a + b = 3/1. X_{I,1}. X_{J,1} + \pi 4/0.5. X_{I,1}. X_{J,2} + \pi 4/0.8. X_{I,2}. X_{J,1} + \pi 5/0.5. X_{I,2}. X_{J,2}$$

«  $a + b$  » a donc trois valeurs possibles : 3, 4 ou 5 avec pour possibilités respectives : 1, 0.8 et 0.5.

(3) Dans le cas où  $\times_{\pi}$  est l'opérateur algébrique multiplication (« \* »), on obtient :

$$a * b = 2/1. X_{I,1}. X_{J,1} + \pi 4/0.5. X_{I,1}. X_{J,2} + \pi 3/0.8. X_{I,2}. X_{J,1} + \pi 6/0.5. X_{I,2}. X_{J,2}$$

«  $a * b$  » a donc quatre valeurs possibles : 2, 3, 4 ou 6 avec pour possibilités respectives : 1, 0.8, 0.5 et 0.5.

(4) Si deux opérandes sont exprimés dans la même base, i.e. si leurs possibilités sont dépendantes, de sorte que lorsque  $a$  vaut 2 (resp. 3), alors  $b$  vaut 1 (resp. 2) avec le même degré de possibilité.  $a$  et  $b$  ont pour valeurs, dans la base canonique :

- $a = 2/1. X_{I,1} + \pi 3/0.8. X_{I,2}$
- $b = 1/1. X_{I,1} + \pi 2/0.5. X_{I,2}$

Les résultats des calculs donnent :

$$a \times_{\pi} b = [2 \text{ (possibilité 1) ou } 3 \text{ (possibilité 0.8)}] \times_{\pi} [1 \text{ (possibilité 1) ou } 2 \text{ (possibilité 0.8)}]$$

$$a \times_{\pi} b = (2 \times_{\pi} 1) / 1. X_{I,1}. X_{I,1} + \pi (2 \times_{\pi} 2) / 0.8. X_{I,2}. X_{I,1}) \times_{\pi} (3 \times_{\pi} 1) / 0.5. X_{I,1}. X_{I,2} + \pi (3 \times_{\pi} 2) / 0.5. X_{I,2}. X_{I,2}$$

Or,  $X.X = X$  et  $i_1 \neq i_2 \Rightarrow X_{I,i1}. X_{I,i2} = 0_v$ ; on a donc pour résultat :

$$a \times_{\pi} b = (2 \times_{\pi} 1) / 1. X_{I,1} \times (3 \times_{\pi} 2) / 0.5. X_{I,2}$$

Il y a donc seulement deux résultats possibles et non trois. Nous avons inclus la possibilité d'exprimer les sources d'incertitudes liées entre elles.

(5) Dans le cas où  $\times_{\pi}$  est l'opérateur algébrique addition (« + »), on obtient :

$$a + b = 3 / 1. X_{I,1} +_{\pi} 5 / 0.5. X_{I,2}$$

C'est-à-dire :  $a + b = 3$  (possibilité 1) ou 5 (possibilité 0.5).

(6) Dans le cas où  $\times_{\pi}$  est l'opérateur algébrique multiplication (« \* »), on obtient :

$$a * b = 2 / 1. X_{I,1} +_{\pi} 6 / 0.5. X_{I,2}$$

C'est-à-dire :  $a * b = 2$  (possibilité 1) ou 6 (possibilité 0.5).

(7) Autre exemple :

$$a \times a = (2 \times 2) / 1. X_{I,1} +_{\pi} (3 \times 9) / 0.8. X_{I,2}$$

Dans le cas où  $\times$  est l'opérateur algébrique addition (« + »), on obtient :

$$a + a = 4 / 1. X_{I,1} +_{\pi} 5 / 0.8. X_{I,2} = 2 * a$$

Dans le cas où  $\times$  est l'opérateur algébrique multiplication (« \* »), on obtient :

$$a * a = 4 / 1. X_{I,1} +_{\pi} 9 / 0.8. X_{I,2} = a^2$$

### 5.3.3 Quantités probabilistes discrètes

#### 5.3.3.1 Loi de composition interne $+_p$

**Définition :** nous définissons une loi de composition interne sur  $D_p(G)$ , notée  $+_p$ , comme l'union de deux quantités probabilistes formalisée par l'expression suivante :

$$\forall (b_1, b_2) \in G^2, P(b_1 +_p b_2) = P(b_1) +_p P(b_2)$$

**Définition :** nous définissons la forme canonique purement probabiliste  $b$  par l'expression suivante :

$$b = b_1 / \beta_1. X^{I,1} +_p \dots +_p b_n / \beta_n. X^{I,n}$$

C'est-à dire :

$$b = \sum_{j=1}^n b_j / \beta_j. X^{I,j}$$

Avec :

- $b_j$  sont les valeurs probables disjointes de  $b$ .

- $\beta_j$  sont les probabilités associées à chaque valeur probable  $b_j$  (la somme des  $\beta_j$  est égale à 1).
- Par hypothèse  $\forall i, j \in \mathbb{N}, b_i \cap b_j = \emptyset$  et  $b = \sum_{j=1}^n b_j / \beta_j \cdot X^{J,j}$ . Les  $b_j / \beta_j \cdot X^{J,j}$  (J fixé) représentent la partition de l'univers des possibles  $\Omega$  correspondant aux valeurs de  $b$ .
- Contrainte d'additivité :  $\sum_{i=1}^n b_i = 1$

L'écriture de  $a$  est unique, les  $b_j$  (resp  $\beta_j$ ) sont dits valeurs canoniques (respectivement poids canonique de  $b$ ).  $b$  s'interprète comme une valeur susceptible de prendre la valeur  $b_j$  avec la probabilité associée  $\beta_j$ . Il s'agit ici typiquement d'une disjonction discrète et finie d'éléments de  $G$ , chacun étant associé à une probabilité. Autrement dit, il s'agit de variables probabilistes sur  $\Omega$ , prenant un nombre fini de valeurs  $(b_i)_{i=1}^n \in G^n$ .

*Propriétés de  $(D_p(G), +_p)$*

- $0_v$  est l'élément neutre pour  $+_p$ .
- $+_p$  est commutative et associative.
- Il n'existe pas d'inverse pour  $+_p$ .
- $(D_p(G), +_p)$  est un demi-groupe unitaire (ou monoïde) commutatif.

*Exemples*

(1) Base de données orientée objet probabiliste : la notation ci-dessous correspond à une forme développée de base de données orientée objet probabiliste à laquelle on a associé des vecteurs de la base de  $E$ .

$$B = \sum_{i=1}^n B_i / P_i \cdot X_{J,i} \text{ avec } \sum_{i=1}^n P_i = 1 \text{ et } B_i = \prod_{k=1}^K \langle o_{i,k}, \overrightarrow{v_{i,k}} \rangle \text{ où } B_i \text{ est une base nette}$$

(2) Soient  $a$  et  $b$  deux nombres qui ont plusieurs valeurs probables dans l'ensemble  $\mathbb{R}$ .

- $b = 2$  (probabilité 0.7) ou 3 (probabilité 0.3)

$b$  s'exprime dans la base canonique avec la formule suivante, avec  $I \neq J$  :

- $b = 2/0.7 \cdot X^{J,1} + 3/0.3 \cdot X^{J,2}$

### 5.3.3.2 Loi de composition interne $\times_p$

**Définition :** soit  $(b_1, b_2) \in D_p(G)^2$  deux quantités purement probabilistes indépendantes  $b_1 = \sum_{i=1}^{n_1} b_{1,i}/\beta_{1,i} \cdot X^{1,i}$  et  $b_2 = \sum_{j=1}^{n_2} b_{2,j}/\beta_{2,j} \cdot X^{2,j}$ . Nous définissons une loi de composition interne sur  $D_p(G)$  notée  $\times_p$  formalisée par l'expression suivante :

$$\forall (b, b_1, b_2) \in G^3, \forall b \in P(C), \beta_{i,j}(b_1 \times_p b_2) = \beta_i(b_1) * \beta_j(b_2)$$

Par hypothèse on a :  $\forall i, j \in \mathbb{N}, b_i \cap b_j = \emptyset$ . Les quantités agrégées sont donc indépendantes, c'est-à-dire que quelles que soient les valeurs  $(b_{1,i})_{1 \leq i \leq n_1}$  et  $(b_{2,j})_{1 \leq j \leq n_2}$ , on a :  $P((b_1 = b_{1,i}) \cap (b_2 = b_{2,j})) = P(b_1 = b_{1,i}) * P(b_2 = b_{2,j})$ .

Remarque : la démarche dans le cas probabiliste est analogue au cas possibiliste. La principale différence réside dans le fait que les probabilités s'agrègent par des « produits ».

L'expression suivante :

$$b_1 \times_p b_2 = (b_{1,1}/\beta_{1,1} \cdot X^{1,1} +_p \dots +_p b_{1,n_1}/\beta_{1,n_1} \cdot X^{1,n_1}) \times_p (b_{2,1}/\beta_{2,1} \cdot X^{2,1} +_p \dots +_p b_{2,n_2}/\beta_{2,n_2} \cdot X^{2,n_2})$$

est évaluée avec une somme composée des  $(b_{1,i} \times_p b_{2,j})/(\beta_{1,i} * \beta_{2,j}) \cdot X^{1,i} \cdot X^{2,j}$

*Démonstration*

$$\begin{aligned} b_1 \times_p b_2 &= \sum_{i=1}^{n_1} b_{1,i}/\beta_{1,i} \cdot X^{1,i} \times_p \sum_{j=1}^{n_2} b_{2,j}/\beta_{2,j} \cdot X^{2,j} \\ b_1 \times_p b_2 &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (b_{1,i}/\beta_{1,i} \cdot X^{1,i} \times_p b_{2,j}/\beta_{2,j} \cdot X^{2,j}) \end{aligned}$$

Les événements «  $b_{1,i}$  » et «  $b_{2,j}$  » étant par hypothèse indépendants deux à deux. On obtient :

$$b_1 \times_p b_2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (b_{1,i} \times_p b_{2,j})/(\beta_{1,i} * \beta_{2,j}) \cdot X^{1,i} \cdot X^{2,j}$$

Ce type de structure correspond au produit cartésien d'éléments de l'ensemble  $G$  (sous forme de disjonction de conjonction de valeurs de  $G$ ), chacun étant affecté à une probabilité. Si  $b_1$  et  $b$  sont compatibles algébriquement, on pourra agréger leurs valeurs avec des lois de

compositions internes telles que + ou \*.  $D_p(G)$  héritera ainsi de certaines propriétés de G pour les lois + et \*. On vérifie que l'on peut immerger l'ensemble G dans  $D_p(G)$ .

### *Généralisation*

On déduit immédiatement la formule suivante pour n combinaisons d'éléments de  $G^n$  :

$$b_1 \times_p \dots \times_p b_n = \left( \prod_{k=1}^n b_i \right) = \sum_{i_1=1}^{N_1} \dots \sum_{i_n=1}^{N_n} \left( \prod_{k=1}^n b_{k,i_k} \right) / \left( \prod_{p=1}^n \alpha_{p,i_p} \right) \cdot \left( \prod_{q=1}^n X^{q,i_q} \right)$$

NB : les grands symboles produits n'ont pas la même signification ( $\times_p$ , \* et .) ; c'est pourquoi nous avons séparé les produits.

### *Expression générale d'une distribution probabiliste sur G*

D'une manière générale, en considérant que chaque terme  $a_i$  est une agrégation d'éléments de G par la loi de composition interne  $\times_p$ , tout élément de  $D_p(G)$  est une expression de la forme :

$$b = \sum_{i=1}^n b_i / (\beta_i \cdot \prod_{j=1}^{n_2} X_{j,i})$$

### *Cas particulier de deux opérandes exprimés dans la même base*

Si les deux opérandes sont exprimés sur la même base (c'est-à-dire que leurs probabilités sont dépendantes), alors la formule se simplifie. En effet, intuitivement, a et b sont issus de la même source d'incertitude, i.e. lorsque le premier opérande vaut  $a_1$  (respectivement  $b_1$ ), la seconde vaut  $b_1$  (respectivement  $b_2$ ) avec le même degré de probabilité.

On pose :  $b_1 = \sum_{i=1}^{n_1} b_{1,i} / \beta_{1,i} \cdot X^{1,i}$  et  $b_2 = \sum_{j=1}^{n_1} b_{2,j} / \beta_{1,j} \cdot X^{1,j}$

Remarque : cela revient au cas général mais avec des probabilités et des bases identiques : quelque soit i entier naturel,  $X^{1,i} = X^{2,j}$ ,  $\beta_{1,i} = \beta_{2,i}$  et  $n_1 = n_2$ . On a alors la situation suivante : lorsque  $b_1$  vaut  $b_{1,i}$ , alors  $b_2$  vaut  $b_{2,j}$ .

En d'autres termes,  $P(b_1 = b_{1,i} / b_2 = b_{2,j}) = 1$  et  $P(b_2 = b_{2,j} / b_1 = b_{1,i}) = 1$ . Avec la formule de Bayes, on en déduit que  $\beta_{1,i} = \beta_{2,i}$  (NB :  $P(b_1 = b_{1,j}) = \beta_{1,j}$  et  $P(b_2 = b_{2,j}) = \beta_{2,j}$ ) :

On a :  $P(b_1 = b_{1,i} \cap b_2 = b_{2,j}) = P(b_1 = b_{1,i} / b_2 = b_{2,j}) \cdot P(b_2 = b_{2,j}) = 1 \cdot P(b_2 = b_{2,j})$

Et :  $P(b_1 = b_{1,i} \cap b_2 = b_{2,j}) = P(b_2 = b_{2,j}/b_1 = b_{1,i}) \cdot P(b_1 = b_{1,i}) = 1 \cdot P(b_1 = b_{1,i})$

Donc :  $P(b_2 = b_{2,j}) = P(b_1 = b_{1,i})$

On a alors :

$$b_1 \times_p b_2 = \sum_{i=1}^{n_1} b_{1,i}/\beta_{1,i} \cdot X^{1,i} \times_p \sum_{j=1}^{n_1} b_{2,j}/\beta_{1,j} \cdot X^{1,j}$$

$$b_1 \times_p b_2 = \sum_{i=1}^{n_1} (b_{1,i}/\beta_{1,i} \cdot X^{1,i} \times_p b_{1,j}/\beta_{1,j} \cdot X^{1,j})$$

$$b_1 \times_p b_2 = \sum_{i=1}^{n_1} (b_{1,i} \times_p b_{2,j})/\beta_{1,i} \cdot X^{1,i}$$

Ce cas particulier prend en compte rigoureusement les dépendances entre sources d'incertitudes afin de ne pas augmenter de manière artificielle le nombre de valeurs probables.

*Propriétés de  $(D_p(G), \times_p)$*

$(D_p(G), \times_p)$  possède les mêmes propriétés que  $(D_\pi(G), \times_\pi)$ , les valeurs possibles (chacune associée à un degré de possibilité) étant remplacées par des valeurs probables (chacune associée à un degré de probabilité).

Supposons que  $(G, +)$  et  $(G, *)$  sont des demi-groupes commutatifs unitaires, i.e.  $G$  possède deux éléments neutres pour  $+$  et  $*$  respectivement notés  $0_G$  et  $1_G$ .

On note les éléments neutres pour  $(D_p(G), +)$  et  $(D_p(G), *)$  respectivement  $0_{D_p(G)}$  et  $1_{D_p(G)}$ .

Les propriétés essentielles sont les suivantes :

- $(D_p(G), +, *)$  possède une structure d'anneau commutatif unitaire.
- Si  $K$  est un corps,  $(D_p(K), +, \cdot)$  avec  $\cdot$  produit externe d'une distribution probabiliste d'éléments d'un corps  $K$  par un élément d'un corps  $K$  (loi de composition de  $K \times D_p(K) \rightarrow D_p(K)$ ) possède une structure d'espace vectoriel sur  $K$  ( $K$  peut par exemple être  $\mathbb{R}$ ).

*Exemples*

(1) Soient  $a$  et  $b$  deux nombres qui ont plusieurs valeurs probables dans l'ensemble  $\mathbb{R}$ .

- $a = 2$  (probabilité 0.7) ou 3 (probabilité 0.3)
- $b = 1$  (probabilité 0.5) ou 2 (probabilité 0.5)

$a$  et  $b$  s'expriment dans la base canonique avec les formules suivantes, avec  $I \neq J$  :

- $a = 2/0.7 \cdot X^{I,1} +_p 3/0.3 \cdot X^{I,2}$
- $b = 1/0.5 \cdot X^{J,1} +_p 2/0.5 \cdot X^{J,2}$

En appliquant les résultats obtenus précédemment, on obtient :

$$a \times_p b = (2 / 0.7 \cdot X^{I,1} +_p 3 / 0.3 \cdot X^{I,2}) \times_p (1/0.5 \cdot X^{J,1} +_p 2/0.5 \cdot X^{J,2})$$

$$a \times_p b = (2 \times_p 1) / 0.35 \cdot X^{I,1} \cdot X^{J,1} +_p (3 \times_p 1) / 0.15 \cdot X^{I,2} \cdot X^{J,1} +_p (2 \times_p 2) / 0.35 \cdot X^{I,1} \cdot X^{J,2} +_p (3 \times_p 2) / 0.15 \cdot X^{I,2} \cdot X^{J,2}$$

(2) Dans le cas où  $\times_p$  est l'opérateur algébrique addition (« + »), on obtient :

$$a + b = 3 / 0.35 \cdot X^{I,1} \cdot X^{J,1} +_p 4 / 0.15 \cdot X^{I,2} \cdot X^{J,1} +_p 4 / 0.35 \cdot X^{I,1} \cdot X^{J,2} +_p 5 / 0.15 \cdot X^{I,2} \cdot X^{J,2}$$

«  $a + b$  » a donc trois valeurs probables : 3, 4 ou 5 avec pour probabilités respectives : 0.35, 0.5 et 0.15.

(3) Dans le cas où  $\times_p$  est l'opérateur algébrique multiplication (« \* »), on obtient :

$$a * b = 2 / 0.35 \cdot X^{I,1} \cdot X^{J,1} +_p 3 / 0.15 \cdot X^{I,2} \cdot X^{J,1} +_p 4 / 0.35 \cdot X^{I,1} \cdot X^{J,2} +_p 6 / 0.15 \cdot X^{I,2} \cdot X^{J,2}$$

«  $a * b$  » a donc quatre valeurs probables : 2, 3, 4 ou 6 avec pour probabilités respectives : 0.35, 0.15, 0.35 et 0.15.

(4) Le cas de valeurs à probabilités dépendantes s'exprime dans la même base et revient à exprimer des valeurs conditionnelles. Exemple :  $a$  vaut 2 (probabilité 0.7) lorsque  $b$  vaut 1 et  $a$  vaut 3 (probabilité 0.3) lorsque  $b$  vaut 2.  $a$  et  $b$  sont alors exprimées avec la base canonique de la manière suivante :

- $a = 2/0.7 \cdot X^{I,1} +_p 3/0.3 \cdot X^{I,2}$
- $b = 1/0.7 \cdot X^{I,1} +_p 2/0.3 \cdot X^{I,2}$

NB : les probabilités conditionnelles sont alors :  $P(a=2/b=1) = 1$  ;  $P(b=1/a=2)=1$  ;  $P(a=3/b=2) = 1$  ;  $P(b=2/a=3)=1$ .

Or,  $X \cdot X = X$  et  $i_1 \neq i_2 \Rightarrow X^{I,i_1} \cdot X^{I,i_2} = 0_v$  ; on a donc pour résultat :

$$a \times_p b = (2 \times_p 1) / (0.7) \cdot X^{I,1} +_p (2 \times_p 3) / (0.3) \cdot X^{I,2}$$

Les cas  $(2 \times_p 1)$  d'une part et  $(2 \times_p 3)$  d'autre part forment une partition des résultats possibles. La somme de leurs probabilités est égale à 1.



(5) Dans le cas où  $\times_p$  est l'opérateur algébrique addition (« + »), on obtient :

$$a + b = 3 / (0.7) \cdot X^{J,1} +_p 5 / (0.3) \cdot X^{J,2}$$

C'est-à-dire :  $a + b = 3$  (probabilité 0.7) ou 5 (probabilité 0.3).

(6) Dans le cas où  $\times_p$  est l'opérateur algébrique multiplication (« \* »), on obtient :

$$a * b = 2 / (0.7) \cdot X^{J,1} +_p 6 / (0.3) \cdot X^{J,2}$$

C'est-à-dire :  $a * b = 2$  (probabilité 0.7) ou 6 (probabilité 0.3).

(7) Autre exemple :

$$a \times_p a = (2 \times 2) / 0.7 \cdot X^{J,1} +_p (3 \times 3) / 0.3 \cdot X^{J,2}$$

Dans le cas où  $\times_p$  est l'opérateur algébrique addition (« + »), on obtient :

$$a + a = 4 / 0.7 \cdot X^{J,1} +_p 6 / 0.3 \cdot X^{J,2} = 2 * a$$

Dans le cas où  $\times_p$  est l'opérateur algébrique multiplication (« \* »), on obtient :

$$a * a = 4 / 0.7 \cdot X^{J,1} +_p 9 / 0.3 \cdot X^{J,2} = a^2$$

Cela signifie que l'on n'a réalisé qu'un seul tirage de  $a$  dans une loterie.

### 5.3.4 Quantités mixtes discrètes

#### 5.3.4.1 Loi de composition interne $\times$

Rappel :  $D_I(G)^{31} = D_p(G) \cup D_\pi(G)$

Soit  $a = \sum_{i=1}^{n_1} a_i / \alpha_i \cdot X_{I,i}$  une quantité purement possibiliste.

Soit  $b = \sum_{j=1}^{n_2} b_j / \beta_j \cdot X^{J,j}$  une quantité purement probabiliste.

**Définition :** nous définissons une loi de composition interne sur  $D_I(G)$  notée  $\times$ , à savoir :

$$a \times b = (a_{1,1} / \alpha_{1,1} \cdot X_{I,1} +_\pi \dots +_\pi a_{1,n_1} / \alpha_{1,n_1} \cdot X_{I,n_1}) \times (b_{2,1} / \beta_{2,1} \cdot X^{J,1} +_p \dots +_p b_{2,n_2} / \beta_{2,n_2} \cdot X^{J,n_2})$$

**C'est-à-dire, plus formellement :**

$$a \times b = \left( \sum_{i=1}^{n_1} a_i / \alpha_i \cdot X_{I,i} \right) \times \left( \sum_{j=1}^{n_2} b_j / \beta_j \cdot X^{J,j} \right)$$

---

<sup>31</sup>  $D_I(G)$  : distribution incertaine sur  $G$

L'expression ainsi définie est un produit cartésien à valeurs dans  $D_\pi(G) \times D_p(G)$ .

Dans ce cas, il n'est pas possible de distribuer les valeurs possibilistes sur les valeurs probabilistes sans perdre d'information concernant les incertitudes car cela nécessite d'interpréter d'une manière particulière les combinaisons de valeurs probabilistes et possibilistes. Ainsi, si on voulait interpréter les distributions de possibilité comme des mesures de Demspter-Shafer (ce qui implique de transférer une partie des valeurs de possibilités à des masses), les éléments focaux résultants ne seraient pas calculables de manière triviale, la mémoire de la distribution de possibilité originale étant à ce moment-là altérée par le fait d'avoir distribué les valeurs possibilistes sur les valeurs probabilistes.

### ***Recombinaisons avec d'autres éléments de $D_I(G)$***

Dans la suite de ce paragraphe, nous faisons l'hypothèse d'indépendance entre connaissances probabiliste et possibiliste. Les combinaisons croisées probabilités / possibilités restent telles quelles jusqu'à la fin des calculs, i.e. jusqu'à ce que les combinaisons entre quantités possibilistes d'une part et probabilistes d'autre part soient terminées.

En pratique, on réalise les calculs rigoureusement et indépendamment sur les deux espaces probabiliste et possibiliste. Pour combiner les deux formes de connaissances, il faut ensuite fournir une interprétation particulière (i.e. une évaluation) pour évaluer la combinaison entre les connaissances probabiliste et possibiliste (voir chapitre suivant).

**Propriété : la loi  $\times$  est commutative. En effet, il s'agit d'une conjonction divisant l'univers des possibles, sans traiter les disjonctions de valeurs internes.**

Exemple :

$$a \times b \times c = \left( \sum_{i=1}^{n_1} a_i / \alpha_i \cdot X_{I,i} \right) \times \left( \sum_{j=1}^{n_2} b_j / \beta_j \cdot X^{I,j} \right) \times \left( \sum_{k=1}^{n_3} c_k / \gamma_k \cdot X_{K,k} \right)$$

$a \times b$  divise l'univers des possibles par combinatoire sans traiter les disjonctions de valeurs internes.

$$a \times b \times c = a \times c \times b = \left( \sum_{i=1}^{n_1} a_i / \alpha_i \cdot X_{I,i} \right) \times \left( \sum_{k=1}^{n_3} c_k / \gamma_k \cdot X_{K,k} \right) \times \left( \sum_{j=1}^{n_2} b_j / \beta_j \cdot X^{I,j} \right)$$

$a \times b$  constitue une partition de l'univers des possibles. On réalise les combinaisons par les opérateurs classiques de la logique floue.

$$a \times b \times c = \left( \sum_{i=1}^{n_1} \sum_{k=1}^{n_3} (a_i \times c_k) / (\alpha_i \min \gamma_k) \cdot X_{1,i} \cdot X_{3,k} \right) \times \left( \sum_{j=1}^{n_2} b_j / \beta_j \cdot X^{2,j} \right)$$

De la même manière, on peut combiner l'expression précédente avec une distribution purement probabiliste  $d = \sum_{l=1}^{n_4} d_l / \delta_l \cdot X^{4,l}$ , en considérant que  $c \times d$  constitue une partition de l'univers dans l'espace des probabilités.

$$a \times b \times c \times d = \left( \sum_{i=1}^{n_1} a_i / \alpha_i \cdot X_{1,i} \right) \times \left( \sum_{k=1}^{n_3} c_k / \gamma_k \cdot X_{3,k} \right) \times \left( \sum_{j=1}^{n_2} b_j / \beta_j \cdot X^{2,j} \right) \\ \times \left( \sum_{l=1}^{n_4} d_l / \delta_l \cdot X^{4,l} \right)$$

$$a \times b \times c \times d = \left( \sum_{i=1}^{n_1} \sum_{k=1}^{n_3} (a_i \times c_k) / (\alpha_i \min \gamma_k) \cdot X_{1,i} \cdot X_{3,k} \right) \\ \times \left( \sum_{j=1}^{n_2} \sum_{l=1}^{n_4} (b_j \times d_l) / (\beta_j * \delta_l) \cdot X^{2,j} \cdot X^{4,l} \right)$$

La loi de composition  $\times$  correspond respectivement aux opérateurs  $\times_\pi$  dans  $D_\pi(G)$  et  $\times_p$  dans  $D_p(G)$ . Elle permet de combiner entre elles les quantités possibilistes d'une part et les quantités probabilistes d'autre part.

L'expression ainsi définie constitue un produit cartésien à valeurs dans  $(D_\pi(G) \times D_p(G))$ . L'étape suivante consiste à évaluer la combinaison entre les deux types de connaissances probabiliste et possibiliste. Elle sera abordée dans le chapitre suivant.

### *Expression générale d'une distribution incertaine sur G*

**D'une manière générale, en considérant que chaque terme  $c_i$  est une agrégation d'éléments de G par la loi de composition interne  $\times$ , tout élément de  $D_I(G)$  est une expression de la forme :**

$$a \times b = \left( \sum_{i=1}^{m_1} a_i / (\alpha_i \cdot \prod_{l=1}^{n_1} X_{1,i}) \right) \times \left( \sum_{j=1}^{m_2} b_j / (\beta_j \cdot \prod_{j=1}^{n_2} X_{2,j}) \right)$$

*Propriétés de  $(D_\pi(G) \times D_p(G))$  (produit cartésien)*

De manière triviale,  $(D_\pi(G) \times D_p(G))$  peut être muni des mêmes structures algébriques que  $D_\pi(G)$  et  $D_p(G)$  en définissant les opérations composante par composante de par les propriétés du produit cartésien. Par exemple les éléments neutres sont les produits cartésiens des éléments neutres. Cependant, notre but est de pouvoir définir des opérations algébriques combinant des valeurs des deux univers de  $D_\pi(G)$  et de  $D_p(G)$ . Comme nous l'avons déjà précisé, pour combiner les deux formes de connaissances, il faudra ensuite fournir une interprétation particulière (i.e. une évaluation) de combinaison entre les deux types de connaissances (probabiliste et possibiliste).

#### 5.3.4.2 Structure récursive de l'ensemble $D_I$

Il est également nécessaire d'exprimer des connaissances dont les probabilités et possibilités sont dépendantes entre elles. Par exemple, on veut modéliser que la récolte sera « bonne » (possibilité 1) ou « très bonne » (possibilité 0.3) (car climat propice de probabilité 0.6) ou bien « médiocre » (possibilité 1) ou « mauvaise » (possibilité 0.4) (car mauvais temps de probabilité 0.4). Il s'agit typiquement d'une distribution de probabilité à valeurs dans des ensembles flous. Dans ce cas, il n'y a pas indépendance entre les connaissances possibiliste et probabiliste. Les distributions de ce type sont non nécessairement décomposables.

$a = [\text{« bonne » (possibilité 1) ou « très bonne » (possibilité 0.3)}] \text{ (probabilité 0.6) ou } [\text{« médiocre » (possibilité 1) ou « mauvaise » (possibilité 0.5)}] \text{ (probabilité 0.4)}$

a peut être exprimé suivant le formalisme défini précédemment :

$$a = [\text{« bonne »} / 1.X_{I,1} + \text{« très bonne »} / 0.3.X_{I,2}].0.6.X^{J,1} + [\text{« médiocre »} / 1.X_{K,1} + \text{« mauvaise »} / 0.3.X_{K,2}].0.4.X^{J,2}$$

La différence avec le cas précédent est que l'on est en présence de deux demi-univers des possibles : il n'est pas possible d'aller plus loin dans cette représentation sans perdre l'information liée à la dépendance entre probabilités et possibilités. Cette représentation est rigoureuse car il est ensuite possible de recombinaison ce résultat avec d'autres quantités sans confondre les deux demi-univers des possibles.

Les éléments d'un tel ensemble forment une structure plus vaste que  $D_I(G)$ . Il s'agit d'une structure récursive où  $G$  peut-être soit un élément de  $D_\pi(G)$  soit un élément de  $D_p(G)$ .

Par la suite nous noterons cette structure  $D_{IR}(G)^{32}$ . Cette structure peut être vue comme un arbre dont les feuilles sont des valeurs de (possibilité, probabilité) associées à un produit de vecteurs de base probabilistes et possibilistes. Les valeurs à l'extrémité de l'arbre (i.e. les feuilles) correspondent aux données brutes qui appartiennent à l'ensemble  $G$ . Il peut par exemple s'agir d'un corps  $K$ .

Nous venons donc de fournir une structure algébrique permettant de formaliser et de réaliser des calculs rigoureux sur des quantités probabilistes et possibilistes discrètes sans perte de connaissance. Nous allons généraliser notre approche à des quantités continues. Ensuite, dans le chapitre suivant, nous fournirons des interprétations particulières (i.e. des évaluations) de combinaisons entre les deux types de connaissances (probabiliste et possibiliste).

## 5.4 Formalisation et calculs génériques sur des connaissances probabilistes et possibilistes continues

Comme nous l'avons précisé dans le chapitre précédent, les quantités en jeu nécessitent de prendre en compte des quantités continues. Par exemple, on peut avoir à modéliser des connaissances du type : « la valeur de cette surface est soit de 2 hectares environ, soit de 3 hectares environ », ou bien « elle est comprise entre 4 et 5 hectares », ou encore « elle suit une loi normale centrée sur 5 et de variance 0.5 ». Nous considérons que l'incertitude sur l'existence d'une valeur (cas discret représentant les états possibles/probables du monde) est indépendant de la variation autour une valeur particulière (qui est une variation continue sous forme d'intervalle, de distribution de probabilité, de possibilité).

Dans la suite de ce chapitre, nous proposerons un formalisme similaire à celui des distributions incertaines discrètes précédemment présentées. Ce formalisme sera adapté aux quantités continues. Sauf mention contraire, les notations et hypothèses utilisées dans ce chapitre sont les mêmes que dans le cas discret traité précédemment.

Nous définissons les expressions mathématiques de valeurs possibles/probables suivantes :

$$\hat{c}_i = c_i \cdot V(\lambda_i)$$

Avec :

---

<sup>32</sup> Distribution incertaine récursive

- $c_i \in \mathbb{R}$  est la valeur principale.
- $\lambda_i$  est l'amplitude de la variation.
- $V$  est une forme de variation de base continue dont la distribution est possibiliste et/ou probabiliste. Dans la suite de ce document, elle sera appelée variateur.

Par hypothèse, nous considérons que d'une part les disjonctions de variateurs possibilistes ne se superposent pas et d'autre part que les superpositions de variateurs probabilistes sont négligeables.

Nous nous limiterons aux types de variateurs suivants :

- Variateur probabiliste : loi normale de moyenne  $c_i$  et de variance  $\sigma_i^2$ , notée  $c_i.N(\sigma_i^2)$  ;
- Variateur possibiliste : trapézoïde symétrique de mode  $c_i$ , de noyau  $c_i \pm v_i$  et de support  $c_i \pm \lambda_i : c_i$ , notée  $c_i.T(v_i, \lambda_i)$ .
- « Dirac », i.e.  $\delta_0 = V(0) = T(0) = N(0) = 1$  (pas de variation). Dans ce cas, il s'agit d'une distribution incertaine discrète.

Remarque : on nomme un variateur de type « trapézoïde » une distribution de possibilité en forme de trapèze dont les pentes ne sont pas nécessairement des droites. En effet, lors de produit entre distribution possibilistes en forme de trapèze, les pentes gauche et droite du résultat « s'arrondissent ».

Remarque : pour les distributions de probabilité et de possibilité unimodales et symétriques telles que les gaussiennes ( $c_i.N(\sigma_i^2)$ ) et les formes possibilistes triangulaires ( $c_i.T(0, \lambda_i)$ ), le mode, la médiane et la moyenne sont confondus.

Remarque : pour les distributions de possibilité trapézoïdales ( $c_i.T(v_i, \lambda_i)$ ) ou rectangulaires ( $c_i.T(\lambda_i, \lambda_i)$ ), et symétriques, la médiane et la moyenne sont confondues. Le mode correspond au support pour les distributions uniformes et rectangulaires et au noyau pour les distributions trapézoïdales.

Soit  $D_{pC}(\mathbb{R})$ <sup>33</sup> l'ensemble engendré par les vecteurs et les variateurs probabilistes.

Soit  $D_{\pi C}(\mathbb{R})$ <sup>34</sup> l'ensemble engendré par les vecteurs et les variateurs possibilistes.

---

<sup>33</sup>  $D_{pC}(\mathbb{R})$  : distribution probabiliste continue sur  $\mathbb{R}$

<sup>34</sup>  $D_{\pi C}(\mathbb{R})$  : distribution possibiliste continue sur  $\mathbb{R}$

Soit  $D_{IC}(\mathbb{R})^{35} = D_{pC}(\mathbb{R}) \cup D_{\pi C}(\mathbb{R})$ . Il s'agit de l'ensemble engendré par les vecteurs et les variateurs probabilistes et possibilistes.

Remarques : les calculs sur des quantités discrètes de  $\mathbb{R}$  sont des cas particuliers de calculs sur des quantités continues avec des variateurs de types « Dirac » :

- $D_p(\mathbb{R}) \subset D_{pC}(\mathbb{R})$
- $D_{\pi}(\mathbb{R}) \subset D_{\pi C}(\mathbb{R})$
- $D_I(\mathbb{R}) \subset D_{IC}(\mathbb{R})$

Pour garder les propriétés algébriques du cas discret et mener de manière formelle et systématique les calculs, nous allons étudier la stabilité des variateurs sur le corps  $\mathbb{R}$  pour les lois somme et produit.

#### 5.4.1 Quantités possibilistes continues

Dans ce cas, on se restreint à des coefficients et à des variateurs tous deux possibilistes. Il s'agit donc de distributions purement possibilistes continues. Soit  $a \in D_{\pi C}(\mathbb{R})$ .  $a$  est de la forme :

$$a = \sum_{i=1}^m \hat{a}_i / \alpha_i \cdot \left( \prod_{l=1}^n X_{l,i} \right) = \sum_{i=1}^m a_i \cdot T(v_i, \lambda_i) / \alpha_i \cdot \left( \prod_{l=1}^n X_{l,i} \right)$$

Avec  $a_i \cdot T(v_i, \lambda_i)$  trapézoïde de mode  $a_i$ , de noyau  $a_i \pm v_i$  et de support  $a_i \pm \lambda_i$ .

Remarques :

- Si  $\forall i \in \mathbb{N}, v_i = \lambda_i$ , alors le variateur possibiliste est un rectangle de support (et noyau)  $a_i \pm \lambda_i$  que nous noterons  $T(v_i, \lambda_i)$  ou bien  $T(\lambda_i)$
- Si  $\forall i \in \mathbb{N}, v_i = 0$ , alors le variateur possibiliste est un triangle de noyau nul et de support  $a_i \pm \lambda_i$  que nous noterons  $T(0, \lambda_i)$
- Si  $\forall i \in \mathbb{N}, v_i = \lambda_i = 0$ , alors on se ramène au cas d'une quantité discrète de valeur  $a_i$  que nous noterons  $T(0)$  ou  $\delta_0$  (« Dirac »)

Ce type de distribution est une disjonction possibiliste de valeurs incertaines (au sens possibiliste du terme). On peut le décrire également comme un type de distributions de

---

<sup>35</sup>  $D_{IC}(\mathbb{R})$  : distribution incertaine continue sur  $\mathbb{R}$

possibilité « multimodale » dans le sens où  $a$  peut prendre un certain nombre de valeurs principales  $a_i$ .

*Calculs avec la loi de composition interne  $\times_\pi$*

Soient  $a_1$  et  $a_2$  deux quantités purement possibilistes :  $a_1 = \sum_{i=1}^{n_1} \hat{a}_{1,i}/\alpha_{1,i} \cdot X_{1,i}$  et  $a_2 = \sum_{j=1}^{n_2} \hat{a}_{2,j}/\alpha_{2,j} \cdot X_{2,j}$  avec :  $\hat{a}_{1,i} = a_{1,i} \cdot T(v_{1i}, \lambda_{1i})$  et  $\hat{a}_{2,j} = a_{2,j} \cdot T(v_{2j}, \lambda_{2j})$

On a :

$$a_1 \times_\pi a_2 = \sum_{i=1}^{n_1} \hat{a}_{1,i}/\alpha_{1,i} \cdot X_{1,i} \times_\pi \sum_{j=1}^{n_2} \hat{a}_{2,j}/\alpha_{2,j} \cdot X_{2,j}$$

$$a_1 \times_\pi a_2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\hat{a}_{1,i} \times_\pi \hat{a}_{2,j})/(\alpha_{1,i} \min \alpha_{2,j}) \cdot X_{1,i} \cdot X_{2,j}$$

Les opérations se réalisent sur l'ensemble  $\mathbb{R}$ , avec les opérations algébriques  $+$ ,  $-$ ,  $*$  et  $/$ . La somme de deux trapèzes est invariante par somme et produit (moyennant une approximation pour ce dernier). Il nous faut donc évaluer l'expression suivante :

$$\hat{a}_{1,i} \times_\pi \hat{a}_{2,j} = a_{1,i} \cdot T(v_{1i}, \lambda_{1i}) \times_\pi a_{2,j} \cdot T(v_{2j}, \lambda_{2j})$$

L'arithmétique de distributions possibilistes trapézoïdales est équivalente à l'arithmétique d'intervalles pour chaque alpha-coupe. La propagation de telles distributions à travers des opérations arithmétiques (via le principe d'extension de Zadeh) ont une invariance beaucoup plus large qu'en théorie des probabilités, où seule la loi normale est invariante par addition. Les valeurs obtenues pour les noyaux et supports résultants des quatre opérations arithmétiques  $+$ ,  $-$ ,  $*$  et  $/$  sont les suivantes :

- Addition (« + ») :  $\hat{a}_{1,i} + \hat{a}_{2,j} = (a_{1,i} + a_{2,j}) \cdot T(v_{1i} + v_{2j}, \lambda_{1,i} + \lambda_{2,j})$

$$a_1 + a_2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (a_{1,i} + a_{2,j}) \cdot T(v_{1i} + v_{2j}, \lambda_{1,i} + \lambda_{2,j})/(\alpha_{1,i} \min \alpha_{2,j}) \cdot X_{1,i} \cdot X_{2,j}$$

- Soustraction (« - ») :  $\hat{a}_{1,i} - \hat{a}_{2,j} = (a_{1,i} - a_{2,j}) \cdot T(v_{1i} + v_{2j}, \lambda_{1,i} + \lambda_{2,j})$

$$a_1 - a_2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (a_{1,i} - a_{2,j}) \cdot T(v_{1i} + v_{2j}, \lambda_{1,i} + \lambda_{2,j})/(\alpha_{1,i} \min \alpha_{2,j}) \cdot X_{1,i} \cdot X_{2,j}$$



- Produit (« \* ») :  $\hat{a}_{1,i} * \hat{a}_{2,j} \approx (a_{1,i} * a_{2,j}).T(a_{2,j} * v_{1,i} + a_{1,i} * v_{2,i}, a_{2,j} * \lambda_{1,i} + a_{1,i} * \lambda_{2,i})$ . Nous utilisons le signe «  $\approx$  » car si le noyau et le support résultants sont exacts, la forme de trapèze n'est pas exactement conservée (les pentes « s'arrondissent »). L'approximation en trapèze est d'autant meilleure que la différence entre noyau et support est petite devant la valeur principale [Bouchon-Meunier, 1993].

$$a_1 * a_2 \approx \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (a_{1,i} * a_{2,j}).T(a_{2,j} * v_{1,i} + a_{1,i} * v_{2,i}, a_{2,j} * \lambda_{1,i} + a_{1,i} * \lambda_{2,i}) \\ / (\alpha_{1,i} \min \alpha_{2,j}).X_{1,i}.X_{2,j}$$

- Division (« / ») :  $\hat{a}_{1,i}/\hat{a}_{2,j} \approx (a_{1,i}/a_{2,j}).T\left(\frac{a_{2,j}*v_{1,i}+a_{1,i}*v_{2,i}}{a_{2,j}^2}, \frac{a_{2,j}*\lambda_{1,i}+a_{1,i}*\lambda_{2,i}}{a_{2,j}^2}\right)$ , sachant que  $\hat{a}_{1,i}/\hat{a}_{2,j}$  n'est pas défini lorsque  $\hat{a}_{2,j}$  prend la valeur 0. La même remarque que pour le produit est valable dans ce cas.

$$a_1/a_2 \approx \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (a_{1,i} / a_{2,j}).T\left(\frac{a_{2,j} * v_{1,i} + a_{1,i} * v_{2,i}}{a_{2,j}^2}, \frac{a_{2,j} * \lambda_{1,i} + a_{1,i} * \lambda_{2,i}}{a_{2,j}^2}\right) \\ / (\alpha_{1,i} \min \alpha_{2,j}).X_{1,i}.X_{2,j}$$

Remarque : pour les opérations algébriques « produit » et « division », nous avons choisi d'approximer le résultat par un trapèze grâce à une approximation tangentielle. Cette approximation est valable moyennant l'hypothèse que l'amplitude des variations de valeurs est faible en regard de ces valeurs. Une autre approximation, dite « sécante », qui ne nécessite pas cette hypothèse [Dubois et Prade, 1980], n'est pas utilisée ici par souci de clarté.

*Propriétés des formes trapézoïdales symétriques (T, +, \*)*

On peut définir une véritable structure algébrique des formes trapézoïdales symétriques de mode  $a_i$ , de noyau  $a_i \pm v_i$  et de support  $a_i \pm \lambda_i$  (i.e.  $a_i.T(v_i, \lambda_i)$ ), qui sont stables pour les lois de compositions + et \*.

- + et \* sont commutatives et associatives.
- \* est distributive sur +.
- L'élément neutre pour +, noté  $\delta_0$ , est la variable possibiliste 0.T(0,0) (viateur « Dirac » en 0).  $\delta_0$  est également l'élément absorbant pour \*.

- L'élément neutre pour \*, noté  $\delta_1$ , est la variable possibiliste  $1.T(0,0)$  (viateur « Dirac » en 1)
- Attention, il n'y a pas d'inverse ni pour + ni pour \* :
  - Si  $a = a.T(v_a, \lambda_a)$ , on définit  $-a = -a.T(v_a, \lambda_a)$ , avec  $a - a = 0.T(2v_a, 2\lambda_a) \neq \delta_0$ .
  - Si  $a = a.T(v_a, \lambda_a)$  avec  $(a, v_a, \lambda_a) \neq (0,0,0)$ , on définit  $\frac{1}{a} = \frac{1}{a}.T\left(\frac{1}{v_a}, \frac{1}{\lambda_a}\right)$ , avec  $\frac{a}{a} = 1.T\left(\frac{a^2+v_a^2}{v_a}, \frac{a^2+\lambda_a^2}{\lambda_a}\right) \neq \delta_1$ .
- Si l'on considère que  $v_a$  et  $\lambda_a$  sont très proches de 0, alors on peut considérer que tout élément de T possède un inverse par +. Dans ce cas,  $(T, +, *)$  possède une structure d'anneau commutatif unitaire.  $(T(\mathbb{R}), +, .)$  avec . le produit externe d'une distribution trapézoïdale par une valeur de  $\mathbb{R}$  (loi de composition de  $\mathbb{R} \times T(\mathbb{R}) \rightarrow T(\mathbb{R})$ ) possède alors une structure d'espace vectoriel sur  $\mathbb{R}$ .

### Généralisation

On déduit immédiatement la formule suivante pour n combinaisons d'éléments de  $\mathbb{R}^n$  :

$$\begin{aligned}
 a_1 \times_{\pi} \dots \times_{\pi} a_n &= \left( \prod_{k=1}^n a_i \right) \\
 &= \sum_{i_1=1}^{N_1} \dots \sum_{i_n=1}^{N_n} \left( \prod_{k=1}^n a_{k,i_k} \right) . T(v_{k,i_k}, \lambda_{k,i_k}) / \min(\alpha_{1,i_1}, \dots, \alpha_{n,i_n}) . \left( \prod_{p=1}^n X_{p,i_p} \right)
 \end{aligned}$$

### Expression générale d'une distribution possibiliste continue sur $\mathbb{R}$

D'une manière générale, en considérant que chaque terme  $a_i$  est une agrégation d'éléments de  $\mathbb{R}$  par la loi de composition interne  $\times_{\pi}$ , tout élément de  $D_{\pi C}(\mathbb{R})$  est une expression de la forme :

$$a = \sum_{i=1}^n a_i . T(v_i, \lambda_i) / (\alpha_i . \prod_{l=1}^{n_1} X_{l,i})$$

### Propriétés de $(D_{\pi C}(\mathbb{R}), \times_{\pi})$

Les différences entre les propriétés de  $D_{\pi C}(\mathbb{R})$  et celles de  $D_{\pi}(\mathbb{R})$  concernent les éléments neutres et les inverses pour + et \*.

- En effet, l'élément neutre pour  $+$  est la variable possibiliste associant à tout élément de  $\mathbb{R}$  la valeur « environ 0 ». En fait, il existe une infinité d'éléments neutres car le support d'un trapézoïde contient une infinité d'éléments sur  $\mathbb{R}$ . On les note indistinctement  $0_{D_{\pi C}(\mathbb{R})}$ . En pratique, on définira un majorant du support de trapézoïde centré sur 0, définissant le seuil à ne pas dépasser pour que l'expression « environ » conserve son sens. Dans ce cas, le support des valeurs en jeu dans les calculs seront « petits » devant les valeurs principales.
- En ce qui concerne l'existence d'un inverse pour  $+$ , la combinaison de  $a$  et  $-a$  entraîne une combinaison de trapézoïdes ne donnant pas 0 mais un trapézoïde centré sur 0 et dont les largeurs du noyau et du support sont doublées, ce qui donne :  $a - a = \sum_{i=1}^n 0.T(2\nu_i, 2\lambda_i)/(\alpha_i \cdot \prod_{l=1}^{n_1} X_{l,i})$ , i.e. l'élément neutre est « environ 0 » pour chaque alternative possibiliste. En pratique, on pourra définir un majorant du support des trapézoïdes de sorte à ce que ceux résultant des calculs, aient un support suffisamment petit par rapport aux valeurs principales pour que l'expression « environ » conserve son sens. Par hypothèse, les noyaux et supports des valeurs en jeu seront donc « petits » devant les valeurs principales.
- Le même raisonnement est valable pour les éléments neutres pour  $*$ , notés  $1_{D_{\pi C}(\mathbb{R})}$ , qui sont les variables possibilistes associant à tout élément de  $\mathbb{R}$  la valeur « environ 1 ».
- La propriété de distributivité de  $*$  sur  $+$  est conservée.
- Tout élément de  $D_{\pi C}(\mathbb{R})$  dont aucune valeur possible est « environ égale à zéro » possède également un inverse par  $*$  :  $a = \sum_{i=1}^n a_i.T(\nu_i, \lambda_i)/(\alpha_i \cdot \prod_{l=1}^{n_1} X_{l,i})$  a pour inverse  $1/a = \sum_{i=1}^n (1/a_i).T(\nu_i, \lambda_i)/(\alpha_i \cdot \prod_{l=1}^{n_1} X_{l,i})$  avec  $(a_i)_{i=0}^{i=n} \neq 0$ . Un tel élément sera dit « nécessairement non nul ». Par contre aucun élément de  $D_{\pi C}(\mathbb{R})$  « possiblement environ égal à zéro » (i.e. dont au moins une valeur possible est environ égale à zéro) n'est inversible par  $*$ . Ainsi  $(D_{\pi C}(\mathbb{R}), +, *)$  ne possède pas une structure de corps.
- $(D_{\pi C}(\mathbb{R}), +, *)$  possède une structure d'anneau commutatif unitaire.
- $(D_{\pi C}(\mathbb{R}), +, .)$  avec  $.$  produit externe d'une distribution possibiliste d'éléments de  $\mathbb{R}$  par un élément  $\mathbb{R}$  (loi de composition de  $\mathbb{R} \times D_{\pi C}(\mathbb{R}) \rightarrow D_{\pi C}(\mathbb{R})$ ) possède une structure d'espace vectoriel sur  $\mathbb{R}$ .

### 5.4.2 Quantités probabilistes continues

Dans ce cas, on se restreint à des coefficients et à des variateurs tous deux probabilistes. Nous travaillons sur des distributions purement probabilistes continues sur l'ensemble  $\mathbb{R}$ . Soit  $b \in D_{PC}(\mathbb{R})$ .  $b$  est de la forme :

$$b = \sum_{j=1}^m \hat{b}_j / (\beta_j \cdot \prod_{J=1}^n X^{J,j}) = \sum_{j=1}^m b_j \cdot V_p(\sigma_j^2) / (\beta_j \cdot \prod_{J=1}^n X^{J,j})$$

Avec  $b_j \cdot V_p(\sigma_j^2)$  variateur d'espérance  $b_j$  et de variance  $\sigma_j^2$ .

Ce type de distribution est une disjonction probabiliste de variables aléatoires (ou VA<sup>36</sup>) à loi normale. On peut le décrire également comme un type de distribution de probabilité « multimodale » dans le sens où  $b$  peut prendre un certain nombre de valeurs principales  $b_j$ .

#### *Cas général*

Nous allons tout d'abord traiter le cas général de VA de loi quelconque. Les opérations se réalisent sur l'ensemble  $\mathbb{R}$ , avec les opérations algébriques  $+$  et  $*$ . Il convient donc d'évaluer l'expression suivante :

$$\hat{b}_{1,i} \times_p \hat{b}_{2,j} = b_{1,i} \cdot V_p(\sigma_{1i}^2) \times_p b_{2,j} \cdot V_p(\sigma_{2i}^2)$$

Ce cas étant général, nous nous restreignons à des calculs des moments d'ordres 1 (espérance) et 2 (variance) des combinaisons des VA suivant une loi quelconque.

NB : contrairement au cas d'une distribution de possibilité de type LR (e.g. trapézoïdes), la plupart des lois de probabilités ne sont pas stables par sommes et produits. Cependant, il est possible de déduire certaines propriétés générales de telles combinaisons de VA, telles que l'espérance et la variance.

Dans ce paragraphe, nous présentons les résultats des lois de compositions  $+$  et  $*$  sur des VA réelles.

- Soient  $X$  une VA à valeurs réelles. Sa moyenne et sa variance sont notées respectivement  $m$  et  $\sigma^2$ .

---

<sup>36</sup> Dans la suite nous privilégierons l'acronyme « VA » pour « variable aléatoire »

- Soient  $(X_1, X_2, \dots, X_n)$  n VA indépendantes deux à deux. Leurs moyennes et variances sont notées respectivement  $(m_1, m_2, \dots, m_n)$  et  $(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ .

### Somme de variables aléatoires indépendantes :

Calcul de l'espérance :

- $E(X_1 + X_2 + \dots + X_n) = m_1 + m_2 + \dots + m_n$  par linéarité
- $E(X_1 + X_2 + \dots + X_n) = \sum_{k=1}^n m_k$
- Cas particulier :  $E(n.X) = n.E(X) = n.m$

Calcul de la variance :

- $V(X_1 + X_2 + \dots + X_n) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$  (en effet, la covariance deux à deux des VA est nulle puisqu'elles sont indépendantes).
- $V(X_1 + X_2 + \dots + X_n) = \sum_{k=1}^n \sigma_k^2$

### Produit de variables aléatoires indépendantes :

Calcul de l'espérance :

- $E(X_1.X_2.\dots.X_n) = m_1.m_2.\dots.m_n$  (par indépendance)
- $E(X_1.X_2.\dots.X_n) = \prod_{k=1}^n m_k$

Calcul de la variance :

- $V(X_1.X_2) = E[(X_1.X_2)^2] - E(X_1.X_2)^2$
- $V(X_1.X_2) = E(X_1^2).E(X_2^2) - E(X_1)^2.E(X_2)^2 = (\sigma_1^2 + m_1^2).(\sigma_2^2 + m_2^2) - m_1^2.m_2^2$

Pour n VA  $X_1, X_2, \dots, X_n$ , en utilisant la formule de König-Huygens ( $E(X^2) = \sigma^2 + m^2$ ), on montre que la relation générale entre la variance du produit de n VA indépendantes et leurs variances et espérances revient à développer une expression polynômiale à n variables :

- $V(X_1.X_2.\dots.X_n) = (\sigma_1^2 + m_1^2)(\sigma_2^2 + m_2^2).\dots.(\sigma_n^2 + m_n^2) - m_1^2.m_2^2.\dots.m_n^2$
- $V(X_1.X_2.\dots.X_n) = \prod_{k=1}^n (\sigma_k^2 + m_k^2) - \prod_{k=1}^n m_k^2$

Soient  $(a, b) \in (D_{PC}(\mathbb{R}))^2$ . Avec la combinaison de n variateurs probabilistes indépendants on obtient des VA de lois indéterminées que nous notons de la façon suivante :

(1) La somme de n VA de moyenne  $m_k$  et de variance  $\sigma_k^2$  est notée  $\sum_{k=1}^n m_k.V_k(\sum_{k=1}^n \sigma_k^2)$ .

Exemple pour la somme  $a + b$  :

$$a + b = \sum_{i=1}^{m_a} a_i.V_i(\rho_i^2)/(\alpha_i.\prod_{j=1}^{n_a} X^{l,j}) + \sum_{j=1}^{m_b} b_j.V'_j(\sigma_j^2)/(\beta_j.\prod_{j=1}^{n_b} X^{j,j})$$

$$a + b = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} (a_i + b_j) \cdot (V_i(\rho_i^2) + V'_j(\sigma_j^2)) / (\alpha_i * \beta_j) \cdot \left( \prod_{J=1}^{n_a} X^{I,i} \right) \left( \prod_{J=1}^{n_b} X^{J,j} \right)$$

$$a + b = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} (a_i + b_j) \cdot V''_{i,j}(\rho_i^2 + \sigma_j^2) / (\alpha_i * \beta_j) \cdot \left( \prod_{J=1}^{n_a} X^{I,i} \right) \left( \prod_{J=1}^{n_b} X^{J,j} \right)$$

(2) Le produit de n VA de moyenne  $m_k$  et de variance  $\sigma_k^2$  est noté  $\prod_{k=1}^n m_k \cdot V_k(\prod_{k=1}^n (\sigma_k^2 + mk^2 - k=1 nmk))$ . Exemple pour le produit  $a * b$ :

$$a * b = \sum_{i=1}^{m_a} a_i \cdot V_i(\rho_i^2) / (\alpha_i \cdot \prod_{J=1}^{n_a} X^{I,i}) * \sum_{j=1}^{m_b} b_j \cdot V'_j(\sigma_j^2) / (\beta_j \cdot \prod_{J=1}^{n_b} X^{J,j})$$

$$a + b = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} (a_i * b_j) \cdot (V_i(\rho_i^2) * V'_j(\sigma_j^2)) / (\alpha_i * \beta_j) \cdot \left( \prod_{J=1}^{n_a} X^{I,i} \right) \left( \prod_{J=1}^{n_b} X^{J,j} \right)$$

$$a + b = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} (a_i * b_j) \cdot V''_{i,j} \left( (\rho_i^2 + a_i^2) + (\sigma_j^2 + b_j^2) - a_i^2 \cdot b_j^2 \right) / (\alpha_i * \beta_j) \cdot \left( \prod_{J=1}^{n_a} X^{I,i} \right) \left( \prod_{J=1}^{n_b} X^{J,j} \right)$$

Remarque : il n'est pas possible de donner la loi des sommes et des produits résultant des  $V_k$ , les lois étant quelconques.

Remarque : L'espérance est une application linéaire, i.e.  $\forall a \in \mathbb{R}, E(a \cdot X) = a \cdot E(X) = a \cdot m$ . Changer l'échelle, i.e. multiplier par n, modifie la variance quadratiquement (i.e.  $\forall a \in \mathbb{R}, V(a \cdot X) = a^2 \cdot V(X) = a^2 \cdot \sigma^2$ ).

*Propriétés de  $(D_{pC}(\mathbb{R}), \times_p)$*

Dans le cas de VA quelconques il n'y a pas de loi de composition interne avec + ou \* car la stabilité des variateurs n'est pas assurée automatiquement en l'état actuel des connaissances mathématiques.

Le cas particulier de variateurs gaussiens permet de retrouver certaines propriétés algébriques observées dans le cas discret. On définit alors une loi de composition interne (somme). En effet, la somme de deux VA indépendantes de lois normales sur  $\mathbb{R}$ , de moyennes  $m_1$  et  $m_2$  et

de variances  $\sigma_1$  et  $\sigma_2$  est elle-même une VA de loi normale (par produit de convolution des deux densités) de moyenne  $m_1 + m_2$  et de variance  $\sigma_1^2 + \sigma_2^2$ .

Soient  $a$  et  $b$  deux éléments de  $D_{IC}(\mathbb{R})$  qui sont des disjonctions de respectivement  $i$  et  $j$  VA munies d'une valeur principale (i.e. la moyenne, resp.  $a_i$  et  $b_j$ ) et d'une variabilité (i.e. la variance, resp.  $\rho_i^2$  et  $\sigma_j^2$ ), pondérées de leur probabilité d'occurrence (resp.  $\alpha_i$  et  $\beta_j$ ). Soient  $f$  et  $g$  leurs densités associées.

On exprime la loi de  $a + b$ , ce qui revient à calculer la probabilité d'occurrence de chaque combinaison  $a_i + b_j$  probable :

$$a + b = \sum_{i=1}^{m_a} a_i \cdot N(\rho_i^2) / (\alpha_i \cdot \prod_{j=1}^{n_a} X^{I,i}) + \sum_{j=1}^{m_b} b_j \cdot N(\sigma_j^2) / (\beta_j \cdot \prod_{j=1}^{n_b} X^{J,j})$$

$$a + b = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} (a_i \cdot N(\rho_i^2) + b_j \cdot N(\sigma_j^2)) / (\alpha_i * \beta_j) \cdot \left( \prod_{j=1}^{n_a} X^{I,i} \right) \left( \prod_{j=1}^{n_b} X^{J,j} \right)$$

Les sommes de VA suivant des lois normales se combinent par convolutions et redonnent des lois normales. Soit  $*$  l'opérateur « produit de convolution ». On a :

$a_i \cdot f + b_j \cdot g = (a_i + b_j)(f * g)$  par bilinéarité de l'opérateur  $*$ . Alors :

$$a + b = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} (a_i + b_j) (N(\rho_i^2) + N(\sigma_j^2)) / (\alpha_i * \beta_j) \cdot \left( \prod_{j=1}^{n_a} X^{I,i} \right) \left( \prod_{j=1}^{n_b} X^{J,j} \right)$$

$$a + b = \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} (a_i + b_j) \cdot N(\rho_i^2 + \sigma_j^2) / (\alpha_i * \beta_j) \cdot \left( \prod_{j=1}^{n_a} X^{I,i} \right) \left( \prod_{j=1}^{n_b} X^{J,j} \right)$$

On a bien un résultat encore de la forme  $\sum_{i=1}^m c_i \cdot N(\sigma_i^2) / (\gamma_i \cdot \prod_{j=1}^n X^{J,i})$

*Généralisation pour n combinaisons d'éléments à variateurs gaussiens*

On déduit la formule suivante pour  $n$  combinaisons d'éléments à variateurs gaussiens de  $\mathbb{R}^n$  avec la loi de composition interne  $+$  :

$$b_1 + \dots + b_n = \left( \prod_{k=1}^n b_i \right) = \sum_{i_1=1}^{N_1} \dots \sum_{i_n=1}^{N_n} \left( \prod_{k=1}^n b_{k,i_k} \right) \cdot N \left( \sum_{k=1}^n \sigma_{i_k}^2 \right) / \left( \prod_{p=1}^n \alpha_{p,i_p} \right) \cdot \left( \prod_{q=1}^n X^{q,i_q} \right)$$

NB : les grands symboles produits n'ont pas la même signification ( $\times_p$ ,  $*$  et  $.$ ) ; c'est pourquoi nous avons séparé les produits.

*Expression générale d'une distribution probabiliste à variateur gaussien sur  $\mathbb{R}$*

D'une manière générale, en considérant que chaque terme  $a_i$  est une agrégation d'éléments de  $\mathbb{R}$  par la loi de composition interne  $\times_p$ , tout élément de  $D_{pC}(G)$  est une expression de la forme :

$$b = \sum_{i=1}^n b_i \cdot N(\sigma_j^2) / (\beta_i \cdot \prod_{j=1}^{n_2} X_{j,i})$$

Si les variateurs sont gaussiens,  $D_{pC}(\mathbb{R})$  possède les propriétés suivantes :

- L'élément neutre pour  $+$  est la variable probabiliste associant à tout élément de  $\mathbb{R}$  une gaussienne centrée sur 0. En fait, il existe une infinité d'éléments neutres car le nombre de variances possibles ( $\in \mathbb{R}$ ) est infini. On les note indistinctement  $0_{D_{pC}(\mathbb{R})}$ . En pratique, on pourra définir un majorant de la variance. Dans ce cas, les variances des valeurs en jeu dans les calculs seront « petites » devant les valeurs principales.
- En ce qui concerne l'existence d'un inverse pour  $+$ , la combinaison de  $a$  et  $-a$  a pour résultat une disjonction de gaussiennes centrées sur 0 et dont les largeurs du noyau et du support sont doublées, ce qui donne :  $a - a = \sum_{i=1}^n 0 \cdot N(2\sigma_i^2) / (\alpha_i^2 \cdot \prod_{l=1}^{n_1} X_{l,i})$ , i.e. l'élément neutre est une disjonction de gaussiennes centrées sur 0. En pratique, on définira un majorant de la variance de sorte à ce que ceux résultant des calculs aient une variance suffisamment petite par rapport aux valeurs principales.

### 5.4.3 Quantités mixtes continues

Soit  $a = \sum_{i=1}^{n_1} \hat{a}_i / \alpha_i \cdot X_{I,i} \in D_{\pi C}(\mathbb{R})$  (purement possibiliste)

Soit  $b = \sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{J,j} \in D_{pC}(\mathbb{R})$  (purement probabiliste)

Nous supposons les connaissances probabilistes et possibilistes indépendantes.

**Définition :** nous définissons une loi de composition interne sur  $D_{IC}(G)$  notée  $\times$ . Le résultat est :

$$a \times b = \sum_{i=1}^{n_1} \hat{a}_i / \alpha_i \cdot X_{I,i} \times \sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{J,j}$$



$$a \times b = \left( \sum_{i=1}^{n_1} a_i \cdot T(v_i, \sigma_i) / \alpha_i \cdot X_{I,i} \right) \times \left( \sum_{j=1}^{n_2} b_j \cdot V_j(\sigma_j^2) / \beta_j \cdot X_{J,j} \right)$$

De la même manière que dans le cas discret, on sépare les deux mondes possibiliste et probabiliste jusqu'à la fin des calculs conduisant à l'évaluation via une interprétation particulière de la combinaison possibilités/probabilités. Nous aurons donc une expression de la forme :

$$a \times b = \left( \sum_{i=1}^{m_1} \hat{a}_i / \alpha_i \cdot \left( \prod_{l=1}^{n_1} X_{l,i} \right) \right) \times \left( \sum_{j=1}^{m_2} \hat{b}_j / \beta_j \cdot \left( \prod_{j=1}^{n_2} X_{j,j} \right) \right)$$

C'est-à-dire :

$$a \times b = \left( \sum_{i=1}^{m_1} a_i \cdot T(v_i, \sigma_i) / \alpha_i \cdot \left( \prod_{l=1}^{n_1} X_{l,i} \right) \right) \times \left( \sum_{j=1}^{m_2} b_j \cdot V_j(\sigma_j^2) / \beta_j \cdot \left( \prod_{j=1}^{n_2} X_{j,j} \right) \right)$$

#### 5.4.4 Cas particuliers

Jusqu'ici nous avons travaillé sur des distributions possibilistes à variateurs également possibilistes d'une part et probabilistes à variateurs purement probabilistes d'autre part. Nous allons étudier les cas particuliers de (1) distributions possibilistes à variateurs probabilistes et de (2) distributions probabilistes à variateurs possibilistes.

##### **Variateurs probabilistes continus pondérés par une possibilité**

Soient  $a_1 = \sum_{i=1}^{n_1} \hat{a}_{1,i} / \alpha_{1,i} \cdot X_{1,i}$  et  $a_2 = \sum_{j=1}^{n_2} \hat{a}_{2,j} / \alpha_{2,j} \cdot X_{2,j}$  avec :  $\hat{a}_{1,i} = a_{1,i} \cdot N(\sigma_{1i}^2)$  et  $\hat{a}_{2,j} = a_{2,j} \cdot T(v_{2j}, \lambda_{2j})$ .  $a_2$  est donc une quantité purement possibiliste tandis que  $a_1$  est une quantité possibiliste avec un variateur probabiliste.

On a :

$$a_1 \times a_2 = \sum_{i=1}^{n_1} \hat{a}_{1,i} / \alpha_{1,i} \cdot X_{1,i} \times \sum_{j=1}^{n_2} \hat{a}_{2,j} / \alpha_{2,j} \cdot X_{2,j}$$

$$a_1 \times a_2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (a_{1,i} \cdot N(\sigma_{1i}^2) \times a_{2,j} \cdot T(v_{2j}, \lambda_{2j})) / (\alpha_{1,i} \cdot \min \alpha_{2,j}) \cdot X_{1,i} X_{2,j}$$

Le calcul de  $a_1 \times a_2$  nécessite une interprétation particulière de la combinaison entre deux variateurs : l'un, possibiliste ( $a_{2,j}.T(v_{2j}, \lambda_{2j})$ ) et l'autre, probabiliste ( $a_{1,i}.N(\sigma_{1i}^2)$ ). De telles interprétations seront proposées dans le chapitre suivant.

### Variateurs possibilistes continus pondérés par une probabilité

Soient  $b_1 = \sum_{i=1}^{m_1} \hat{b}_{1,i}/\beta_{1,i}.X^{1,i}$  et  $b_2 = \sum_{j=1}^{m_2} \hat{b}_{2,j}/\beta_{2,j}.X^{2,j}$  avec :  $\hat{b}_{1,i} = b_{1,i}.T(v_{1i}, \lambda_{1i})$  et  $\hat{b}_{2,j} = b_{2,j}.N(\sigma_{2j}^2)$ .  $b_2$  est donc une quantité purement probabiliste tandis que  $b_1$  est une quantité probabiliste avec un variateur possibiliste.

$$b_1 \times b_2 = \sum_{i=1}^{m_1} \hat{b}_{1,i}/\beta_{1,i}.X^{1,i} \times \sum_{j=1}^{m_2} \hat{b}_{2,j}/\beta_{2,j}.X^{2,j}$$

$$a_1 \times a_2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (b_{1,i}.T(v_{1i}, \lambda_{1i}) \times b_{2,j}.N(\sigma_{2j}^2))/(\beta_{1,i} * \beta_{2,j}).X^{1,i}.X^{2,j}$$

Le calcul de  $b_1 \times b_2$  nécessite une interprétation particulière de la combinaison entre deux variateurs : l'un, possibiliste ( $b_{1,i}.T(v_{1i}, \lambda_{1i})$ ) et l'autre, probabiliste ( $b_{2,j}.N(\sigma_{2j}^2)$ ). De telles interprétations seront proposées dans le chapitre suivant. Avant cela, nous proposerons une structure de données de type base de données orientée objet dans le prochain paragraphe. Dans cette approche, nous distinguerons la structure (l'intention), i.e. le modèle du domaine, des données (l'extension), i.e. les instances de la vie réelle dont les connaissances que nous avons peuvent être imparfaites.

## 5.5 Structure informatique : modélisation de l'imperfection dans des bases de données

Considérons un ensemble  $O$  d'objets d'une base de données. On notera  $P(O)$  l'ensemble des parties de  $O$ .  $O$  étant de cardinal  $|O|$ , le cardinal de l'ensemble des sous-ensembles de  $P(O)$  est égal à  $|P(O)| = 2^{|O|}$ . Tout sous-ensemble de  $P(O)$  sera appelé collection sur  $O$ .

### 5.5.1 Modélisation « nette »

Dans notre approche, nous distinguons (1) l'intention, qui correspond au domaine de valeurs (structure attributaire), i.e. le modèle, les classes, le domaine, etc. de (2) l'extension, qui

correspond à l'ensemble des instances, i.e. les données, les objets, etc. En effet, nous modélisons une structure de données comme une base de données définie par un ensemble :

$$B = \{(S_i, E_i); i = 1, \dots, n\}$$

où  $S_i$  et  $E_i$  définissent respectivement l'intention et l'extension. Les types de structures considérés sont les agrégations de types élémentaires parmi lesquels les entiers, les réels, les types énumérés, les références à d'autres objets.

Une extension  $E_i$  est un ensemble d'instances  $\langle O_i, V_i \rangle$  où :

- $O_i$  est un identifiant d'objet issu d'un ensemble  $O$  (qui est l'ensemble des identifiants d'objets).
- $V_i$  est un vecteur de valeurs du domaine défini sur  $E_i$  qui appartient à l'ensemble  $V$ , i.e. l'ensemble des valeurs. Il s'agit typiquement d'un vecteur de valeurs dans  $D_{IC}(G)$  ou à ses sous-ensembles étudiés précédemment.

Voici un exemple décrivant deux parcelles avec le formalisme venant d'être défini :

$$E = \{(oid = "1", (crop = wheat, area = 10 ha, weather = wet, production = 20 t/ha), oid = "2", (crop = maize, area = 5 ha, weather = dry, production = 10 t/ha))\}$$

On note  $B = \{\langle O_1, V_1 \rangle, \dots, \langle O_n, V_n \rangle\}$  la fonction qui associe à  $O_1$  la valeur  $V_1$ , ..., à  $O_n$  la valeur  $V_n$ , etc.

Remarque : toute base de données orientée objet nette pourra alors se noter sous la forme suivante :

$$B = \langle O_1, V_1 \rangle \times \dots \times \langle O_n, V_n \rangle = \prod_{i=1}^n \langle O_i, V_i \rangle$$

Avec  $\times$  produit cartésien d'éléments de  $E_i$ . L'insertion d'un élément  $\langle O_{n+1}, V_{n+1} \rangle$  dans la base  $B$  se traduira par l'application de l'opérateur  $\times$  :

$$B \leftarrow B \times \langle O_{n+1}, V_{n+1} \rangle$$

## 5.5.2 Modélisation de l'imperfection

Nous voulons rendre compte des types d'imperfections de données des systèmes socio-environnementaux, vues dans le chapitre précédent. Pour ce faire, nous avons besoin de

modéliser l'incertitude au niveau de l'existence même d'un élément de la base, ainsi que l'imprécision des valeurs et l'incertitude aléatoire.

Pour cela, nous représentons les valeurs de la base par des collections d'objets qui peuvent être floues [Pollet et Robidou, 1998] et/ou probabilistes [Dantan et al., 2015 (2)], auxquelles nous intégrons la notion supplémentaire de dépendance des sources grâce aux vecteurs  $X_{I,i}$  et  $X^{J,j}$ . En conservant les notations précédemment définies, on pourra représenter un objet d'existence incertaine par deux types d'expressions canoniques de la forme :

Cas possibiliste :

$$\langle O, V \rangle / a \cdot X_{I,i} +_{\pi} \emptyset / b \cdot X_{I,i+1} \text{ (avec } i \geq 0, (a, b) \in [0,1]^2 \text{ et } b = 0 \text{ si } a \neq 1)$$

$$\langle O, V \rangle / a \cdot X^{J,j} +_{\pi} \emptyset / (1 - a) \cdot X^{I,j+1} \text{ (avec } j \geq 0)$$

Avec  $a = \pi(\langle O, V \rangle)$  pour le doute possibiliste sur l'existence de l'extension  $\langle O, V \rangle$ ,  $b = 1 - N(\langle O, V \rangle)$  (nécessité de l'existence de  $\langle O, V \rangle$ ) et enfin  $a = P(\langle O, V \rangle)$  pour la probabilité sur l'existence de  $\langle O, V \rangle$ .

Nous définissons donc des bases de données où à la fois l'existence des objets et les valeurs sont imparfaites. Les valeurs aléatoires ou imprécises sont représentées par un vecteur de valeurs appartenant à l'ensemble  $D_{IC}(G)$  ou à ses sous-ensembles étudiés précédemment.  $D_{IC}(G)$  est muni des lois de composition internes  $+$  (alternative) et  $\times$  (produit cartésien) définies précédemment. Nous définissons trois types de bases de données « imparfaites » :

(1) Les bases de données possibilistes, composées d'objets dont l'existence est entachée d'incertitudes de type possibiliste. Le contenu d'une telle base de données s'exprime comme une distribution de possibilité sur l'ensemble des états possibles de la base. Conformément aux notations précédentes, on notera une telle base de la manière suivante :

$$B_{\pi} = \sum_{i=1}^n B_i / \pi_i \cdot X_i = \sum_{i=1}^n \left( \prod_{l=1}^{m(i)} \langle O_{l,i}, V_{l,i} \rangle / \pi_{l,i} \cdot X_{l,i} \right)$$

Où  $\pi_i$  est la possibilité que la base  $B_{\pi}$  soit la base nette  $B_i$  et  $\pi_i = \min (\pi_{l,i})_{l=1}^n$ .

On pourra également l'exprimer sous forme « factorisée » comme un « produit » d'alternatives de collections (sous-ensemble de  $P(O)$ ) en utilisant la distributivité de  $\times$  sur  $+$  :

$$B_\pi = \prod_{l=1}^n \sum_{i=1}^{m(l)} (\langle O_{l,i}, V_{l,i} \rangle / \pi_{l,i} \cdot X_{l,i})$$

Lorsque la base se compose d'un « produit » d'éléments plus ou moins possibles ou nécessaires de la forme  $\{O, V\}/a \cdot X_{l,i} +_\pi \emptyset / b \cdot X_{l,i+1}$ , on pourra l'exprimer sous la forme (dite « totalement factorisée ») suivante :

$$B_\pi = \prod_{l=1}^n (\langle O_l, V_l \rangle / \lambda_l \cdot X_{l,1} +_\pi \emptyset / \mu_l \cdot X_{l,2})$$

(2) Les bases de données probabilistes, composées d'objets dont l'existence est aléatoire :

$$B_p = \sum_{i=1}^n B_i / p_i \cdot X^i = \sum_{i=1}^n \left( \prod_{l=1}^{m(i)} \langle O_{l,i}, V_{l,i} \rangle / p_{l,i} \cdot X^{l,i} \right)$$

Où  $p_i$  est la possibilité que la base  $B_p$  soit la base nette  $B_i$  et  $p_i = \prod_{l=1}^n p_{l,i}$ .

On pourra également l'exprimer sous forme « factorisée » comme un « produit » d'alternatives de collections (sous-ensemble de  $P(O)$ ) en utilisant la distributivité de  $\times$  sur  $+$  :

$$B_p = \prod_{l=1}^n \sum_{i=1}^{m(l)} (\langle O_{l,i}, V_{l,i} \rangle / p_{l,i} \cdot X^{l,i})$$

Lorsque la base se compose d'un « produit » d'éléments plus ou moins possibles ou nécessaires de la forme  $\{O, V\}/a \cdot X_{l,i} +_\pi \emptyset / b \cdot X_{l,i+1}$ , on pourra l'exprimer sous la forme (dite « totalement factorisée ») suivante :

$$B_p = \prod_{l=1}^n (\langle O_l, V_l \rangle / p_{O_l} \cdot X^{l,1} +_\pi \emptyset / (1 - p_l) \cdot X^{l,2})$$

(3) Les bases de données mixtes, composées d'objets dont à la fois l'existence et les valeurs sont soit incertaines aléatoires, soit incertaines possibilistes. On exprimera de telles combinaisons via le formalisme défini dans la partie précédente, c'est-à-dire plus précisément :

$$B_l = \left( \sum_{i=1}^n B_i / \pi_i \cdot X_i \right) \times \left( \sum_{j=1}^m B_j / p_j \cdot X^j \right)$$

Remarque : on pourra exprimer une telle base comme le « produit » d’alternatives de collections d’objets plus ou moins possibles ou probables (chaque alternative sera soit possibiliste soit probabiliste) :

$$B_I = \prod_{l=1}^n \sum_{i=1}^{m(l)} \left( \langle O_{l,i}, V_{l,i} \rangle / (\pi_{l,i} | p_{l,i}) \cdot (X_{l,i} | X^{l,i}) \right)$$

### Exemples

(1) On peut représenter une situation possibiliste  $S_\pi$  ayant deux objets possibles  $\{O_1, V_1\}$  et  $\{O_2, V_2\}$  de la manière suivante (cas possibiliste) :

$$S_\pi = (\{O_1, V_1\} / \pi_{O_1} \cdot X_{1,1} +_\pi (1 - N_{O_1}) / \emptyset \cdot X_{1,2}) \times_\pi (\{O_2, V_2\} / \pi_{O_2} \cdot X_{2,1} +_\pi (1 - N_{O_2}) / \emptyset \cdot X_{2,2})$$

La formule développée correspondante est décrite ci-dessous :

$$S_\pi = (\langle O_1, V_1 \rangle \times_\pi \langle O_2, V_2 \rangle) / (\pi_{O_1} \min \pi_{O_2}) \cdot X_{1,1} \cdot X_{2,1} +_\pi \langle O_1, V_1 \rangle / \pi_{O_1} \cdot X_{1,1} \cdot X_{2,2} +_\pi \langle O_2, V_2 \rangle / \pi_{O_2} \cdot X_{1,2} \cdot X_{2,1} +_\pi \emptyset / ((1 - N_{O_1}) \min (1 - N_{O_2})) \cdot X_{1,2} \cdot X_{2,2}$$

La situation  $S_\pi$  présente quatre alternatives : l’un avec  $O_1$  et  $O_2$  aux états  $V_1$  et  $V_2$ , l’autre avec  $O_1$  à l’état  $V_1$  et où  $O_2$  n’existe pas, l’autre avec  $O_2$  à l’état  $V_2$  et où  $O_1$  n’existe pas et enfin une alternative où ni  $O_1$  ni  $O_2$  n’existent.

(2) Une représentation probabiliste factorisée d’une situation  $S_p$  peut être représentée de la manière suivante :

$$S_p = (\langle O_1, V_1 \rangle / p_{O_1} \cdot X^{1,1} +_p \emptyset / (1 - p_{O_1}) \cdot X^{1,2}) \times_p (\langle O_2, V_2 \rangle / p_{O_2} \cdot X^{2,1} +_p \emptyset / (1 - p_{O_2}) \cdot X^{2,2})$$

La formule développée correspondante est décrite ci-dessous :

$$\begin{aligned} S_p = & (\langle O_1, V_1 \rangle \times_p \langle O_2, V_2 \rangle) / (p_{O_1} * p_{O_2}) \cdot X^{1,1} \cdot X^{2,1} \\ & +_p (\langle O_1, V_1 \rangle) / (p_{O_1} * (1 - p_{O_2})) \cdot X^{1,1} \cdot X^{2,2} \\ & +_p (\langle O_2, V_2 \rangle) / (p_{O_2} * (1 - p_{O_1})) \cdot X^{1,2} \cdot X^{2,1} \\ & +_p \emptyset / ((1 - p_{O_1}) * (1 - p_{O_2})) \cdot X^{1,2} \cdot X^{2,2} \end{aligned}$$

De la même manière que la situation  $S_\pi$ , la situation  $S_p$  présente quatre alternatives : l'une avec  $O_1$  et  $O_2$  aux états  $V_1$  et  $V_2$ , l'autre avec  $O_1$  à l'état  $V_1$  et où  $O_2$  n'existe pas, l'autre avec  $O_2$  à l'état  $V_2$  et où  $O_1$  n'existe pas et enfin une alternative où ni  $O_1$  ni  $O_2$  n'existent.

(3) Un exemple de base de données mixte est  $S_\pi \times S_p$ .

### 5.5.3 Opérateur de sélection

Dans ce paragraphe, nous définissons l'opérateur de  $\sigma$ -sélection de l'algèbre relationnelle dans le contexte des bases de données « imparfaites ». L'opérateur de  $\sigma$ -sélection, appliqué à un sous-ensemble de la base, consiste à retourner le sous-ensemble des objets (des tuples dans un contexte relationnel) de la base de données vérifiant une condition donnée.

Dans la suite de ce chapitre, cette condition sera supposée atomique et portant sur les valeurs des attributs des objets de la base. Les résultats présentés dans [Prade et Testemale, 1986] et [Dubois et Prade, 1988] permettent alors d'étendre l'opérateur de  $\sigma$ -sélection pour des conditions plus complexes (conjonctions, disjonctions et négations de conditions atomiques). Ainsi, pour une base classique, l'opération de  $\sigma$ -sélection notée  $\sigma_C$ , où  $C$  représente une condition, est définie par l'application  $\sigma_C$  qui, à une base  $B$ , retourne le sous-ensemble des objets vérifiant la condition  $C$ .

(1) Sur une base nette, l'application  $\sigma_C$  s'exprime de la manière suivante :

$$B = \prod_{i=1}^n \langle O_i, V_i \rangle \rightarrow \sigma_C(B) = \sigma_C \left( \prod_{i=1}^n \langle O_i, V_i \rangle \right) = \prod_{i=1}^n \sigma_C(\langle O_i, V_i \rangle)$$

(2) Sur une base possibiliste, l'application  $\sigma_C$  s'exprime de la manière suivante :

$$B_\pi = \prod_{l=1}^n \sum_{i=1}^{m(l)} (\langle O_{l,i}, V_{l,i} \rangle / \pi_{l,i} \cdot X_{l,i})$$

Soit  $\pi_{l,i}$  la possibilité croisée entre le critère  $C$  et les données de  $B_\pi$ . Nous aurons donc :

$$\begin{aligned} \sigma_C(B_\pi) &= \prod_{l=1}^n \sum_{i=1}^{m(l)} \sigma_C(\langle O_{l,i}, V_{l,i} \rangle / \pi_{l,i} \cdot X_{l,i}) \\ \sigma_C(B_\pi) &= \prod_{l=1}^n \sum_{i=1}^{m(l)} (\langle O_{l,i}, V_{l,i} \rangle / (\pi_{l,i} \min \pi'_{l,i}) \cdot X_{l,i}) \end{aligned}$$

(3) Sur une base probabiliste, l'application  $\sigma_C$  s'exprime de la manière suivante :

$$B_p = \prod_{I=1}^n \sum_{i=1}^{m(I)} (\langle O_{I,i}, V_{I,i} \rangle / p_{O_{I,i}} \cdot X^{I,i})$$

Soit  $p'_{O_{I,i}}$  la probabilité croisée entre le critère C et les données de  $B_p$ .

$$\sigma_C(B_p) = \prod_{I=1}^n \sum_{i=1}^{m(I)} \sigma_C(\langle O_{I,i}, V_{I,i} \rangle / p_{O_{I,i}} \cdot X^{I,i})$$

$$\sigma_C(B_p) = \prod_{I=1}^n \sum_{i=1}^{m(I)} (\langle O_{I,i}, V_{I,i} \rangle / (p'_{O_{I,i}} * p_{O_{I,i}}) \cdot X^{I,i})$$

### Exemples

#### (1) Sélection sur une base possibiliste

Soit  $B_\pi$  une base possibiliste constituée d'informations issues par exemple de différents capteurs.

- Possibilité qu'il existe un objet  $\langle O_{I,1}, V_{I,1} \rangle$  (avec  $\pi = 1$  et  $N = 0.6$ ) - par exemple un engin agricole - proche de la position spatiale  $(x_I, y_I)$  au sein du système agricole.
- Possibilité qu'il existe soit un objet  $\langle O_{J,1}, V_{J,1} \rangle$  (avec  $\pi = 1$  et  $N = 0.7$ ) - par exemple un animal - soit un objet  $\langle O_{J,2}, V_{J,2} \rangle$  (avec  $\pi = 0.3$  et  $N = 0$ ) - par exemple un autre animal - proche de la position spatiale  $(x_J, y_J)$  au sein du système agricole.

Nous supposons que les valeurs imprécises décrites (par exemple une position spatiale, une quantité, etc.) sont représentées par des sous-ensembles flous normalisés ou des distributions de possibilité. La base  $B_\pi$  est alors traduite par la collection suivante sous forme factorisée :

$$B_\pi = (\langle O_{I,1}, V_{I,1} \rangle / 1 \cdot X_{I,1} + \emptyset / 0.4 \cdot X_{I,2}) * (\langle O_{J,1}, V_{J,1} \rangle / 1 \cdot X_{J,1} + \langle O_{J,2}, V_{J,2} \rangle / 0.3 \cdot X_{J,2})$$

Nous définissons  $O_\pi$  comme un objet présent dans le système agricole. Il peut s'agir par exemple d'un bâtiment. La requête constitue à retourner les objets de la base dont l'attribut sélectionné (la position) pourrait croiser possiblement ou nécessairement celui (celle) de l'objet  $O_\pi$ . Cela se traduit par l'application de l'opérateur de sélection  $\sigma_C$ , sachant que les possibilités et nécessités que chaque objet de  $B_\pi$  « croise »  $O_\pi$  sont les suivantes :



Tableau 13 : possibilités et nécessités croisées de la base  $B_\pi$  avec l'objet  $O_\pi$

$B_\pi \cap O_\pi$	$O_{I,1}$	$O_{J,1}$	$O_{J,2}$
Possibilité d'intersection avec la position de $O_\pi$	0.7	1	0.8
Nécessité d'intersection avec la position de $O_\pi$	0	0.5	0

$$B_\pi \cap O_\pi = \left( (\langle O_{I,1}, V_{I,1} \rangle / 0.7 \cdot X_{K,1} + \emptyset / 1 \cdot X_{K,2}) / 1 \cdot X_{I,1} + \emptyset / 0.4 \cdot X_{I,2} \right) \\ * \left( (\langle O_{J,1}, V_{J,1} \rangle / 1 \cdot X_{L,1} + \emptyset / 0.5 \cdot X_{L,2}) / 1 \cdot X_{J,1} \right. \\ \left. + (\langle O_{J,2}, V_{J,2} \rangle / 0.8 \cdot X_{M,1} + \emptyset / 0.5 \cdot X_{M,2}) / 0.3 \cdot X_{J,2} \right)$$

$$B_\pi \cap O = (\langle O_{I,1}, V_{I,1} \rangle / 0.7 \cdot X_{K,1} \cdot X_{I,1} + \emptyset / 1 \cdot X_{K,2} \cdot X_{I,1}) \\ * (\langle O_{J,1}, V_{J,1} \rangle / 1 \cdot X_{L,1} \cdot X_{J,1} + \emptyset / 0.5 \cdot X_{L,2} \cdot X_{J,1} + \langle O_{J,2}, V_{J,2} \rangle / 0.3 \cdot X_{M,1} \cdot X_{J,2})$$

La structure ici obtenue est identique à la structure de départ.

## (2) Sélection sur une base probabiliste

Soit  $B_p$  une base probabiliste constituée d'informations sur la composition de sacs d'engrais, issus d'un fournisseur d'engrais minéraux, qui sont composés de trois ingrédients principaux : l'azote (N), le phosphore (P) et le potassium (K). Nous distinguons les engrais simples (qui ne contiennent qu'un seul élément nutritif : N, P ou K) des engrais composés (qui peuvent en contenir deux ou trois : NP, NK, PK ou encore NPK). Nous définissons donc la composition des sacs d'engrais en fonction de la présence ou non de ces ingrédients, associés à leur quantité. Nous supposons ici que les compositions d'engrais sont uniquement les suivantes : N, P, K, NP, NK.

- Probabilité que chaque sac contienne de l'azote (N)  $\langle N, V_{I,1} \rangle$  (avec  $p = 0.7$ ).
- Probabilité que chaque sac contienne soit du phosphore (P)  $\langle P, V_{J,1} \rangle$  (avec  $p = 0.8$ ) soit du potassium (K)  $\langle K, V_{J,2} \rangle$  (avec  $p = 0.2$ ).

Nous supposons que les valeurs aléatoires décrites (par exemple la quantité ou la proportion de chaque ingrédient, etc.) sont représentées par des distributions de probabilité discrètes ou continues. La base  $B_p$  est alors traduite par la collection suivante sous forme factorisée :

$$B_p = (\langle N, V_{I,1} \rangle / 0.7 \cdot X^{I,1} + \emptyset / 0.3 \cdot X^{I,2}) * (\langle P, V_{J,1} \rangle / 0.8 \cdot X^{J,1} + \langle K, V_{J,2} \rangle / 0.2 \cdot X^{J,2})$$

Nous définissons  $O_p$  comme le dosage d'ingrédients de fertilisant permettant d'atteindre l'objectif de production. La requête constitue à retourner les ingrédients constituant les sacs du fournisseur dont les quantités sont « compatibles » avec une récolte répondant aux objectifs de production. Cela se traduit par l'application de l'opérateur de sélection  $\sigma_C$ , sachant que les probabilités que les quantités d'ingrédients constituant chaque objet de  $B_p$  sont compatibles avec  $O_p$  sont les suivantes :

Tableau 14 : probabilités conjointes de la base  $B_p$  avec l'objet  $O_p$

$B_p \cap O_p$	$N$	$P$	$K$
Probabilité d'atteinte de l'objectif de production avec la dose	0.8	0.7	0.2
Probabilité de non atteinte de l'objectif de production avec la dose	0.2	0.3	0.8

$$B_p = (\langle N, V_{I,1} \rangle / 0.7 . X^{I,1} + \emptyset / 0.3 . X^{I,2}) * (\langle P, V_{J,1} \rangle / 0.8 . X^{J,1} + \langle K, V_{J,2} \rangle / 0.2 . X^{J,2})$$

$$B_p \cap O_p = \left( (\langle N, V_{I,1} \rangle / 0.8 . X^{K,1} + \emptyset / 0.2 . X^{K,2}) / 0.7 . X^{I,1} + \emptyset / 0.3 . X^{I,2} \right) \\ * \left( (\langle P, V_{J,1} \rangle / 0.7 . X^{L,1} + \emptyset / 0.3 . X^{L,2}) / 0.8 . X^{J,1} \right. \\ \left. + (\langle K, V_{J,2} \rangle / 0.2 . X^{M,1} + \emptyset / 0.8 . X^{M,2}) / 0.2 . X^{J,2} \right)$$

$$B_p \cap O_p = (\langle N, V_{I,1} \rangle / 0.56 . X^{K,1} . X^{I,1} + \emptyset / 0.14 . X^{K,2} . X^{I,1} + \emptyset / 0.3 . X^{I,2}) \\ * (\langle P, V_{J,1} \rangle / 0.4 . X^{L,1} . X^{J,1} + \emptyset / 0.4 . X^{L,2} . X^{J,1} + \langle K, V_{J,2} \rangle / 0.2 . X^{M,1} . X^{J,2} \\ + \emptyset / 0.8 . X^{M,2} . X^{J,2})$$

La structure ici obtenue est identique à la structure de départ.

Interprétation : on obtient une base de données des ingrédients des sacs d'engrais (azote ou rien et phosphore ou potassium). Chaque ingrédient est associé à la probabilité d'une production répondant aux objectifs en fonction de sa quantité probable (degré de compatibilité présence et quantité de l'ingrédient dans le sac / objectif de production).

#### 5.5.4 Opérateurs d'agrégation

Nous allons étendre les opérateurs d'agrégation classiques suivants à l'existence hypothétique des différents objets qui composent la base de données : AVG, SUM, COUNT, MIN, MAX,

STDDEV, VARIANCE. Dans la suite de ce chapitre, OP désignera un de ces opérateurs d'agrégation.

Soit la base possibiliste  $B_\pi$  dans sa forme développée :

$$B_\pi = (\{O_1, V_1\} \times_\pi \{O_2, V_2\}) / (\pi_{O_1} \min \pi_{O_2}) \cdot X_{1,1} \cdot X_{2,1} +_\pi \{O_1, V_1\} / \pi_{O_1} \cdot X_{1,1} \cdot X_{2,2} +_\pi \{O_2, V_2\} / \pi_{O_2} \cdot X_{1,2} \cdot X_{2,1} +_\pi \emptyset / ((1 - N_{O_1}) \min(1 - N_{O_2})) \cdot X_{1,2} \cdot X_{2,2}$$

Soit la base probabiliste  $S_p$  dans sa forme développée :

$$\begin{aligned} B_p &= (\{O_1, V_1\} \times_p \{O_2, V_2\}) / (p_{O_1} * p_{O_2}) \cdot X^{1,1} \cdot X^{2,1} \\ &+_p (\{O_1, V_1\}) / (p_{O_1} * (1 - p_{O_2})) \cdot X^{1,1} \cdot X^{2,2} \\ &+_p (\{O_2, V_2\}) / (p_{O_2} * (1 - p_{O_1})) \cdot X^{1,2} \cdot X^{2,1} \\ &+_p \emptyset / ((1 - p_{O_1}) * (1 - p_{O_2})) \cdot X^{1,2} \cdot X^{2,2} \end{aligned}$$

Si l'on applique l'opérateur OP aux valeurs respectives  $V_1$  et  $V_2$  (qui sont supposées « agrégeables ») de ces bases de données, on obtient dans les deux cas possibiliste et probabiliste quatre mondes possibles : l'un où  $O_1$  et  $O_2$  existent simultanément, un autre où seul  $O_1$  existe, un autre où seul  $O_2$  existe et enfin l'un où ni  $O_1$  ni  $O_2$  n'existent. Nous obtenons la valeur suivante :

$$\begin{aligned} OP(V_1, V_2) &= OP(V_1, V_2) / (\pi_{O_1} \min \pi_{O_2}) + OP(V_1) / \pi_{O_1} + OP(V_2) / \pi_{O_2} \\ &+ OP(\emptyset) / ((1 - N_{O_1}) \min(1 - N_{O_2})) \end{aligned}$$

$$\begin{aligned} OP(V_1, V_2) &= OP(V_1, V_2) / (p_{O_1} * p_{O_2}) + OP(V_1) / (p_{O_1} * (1 - p_{O_2})) \\ &+ OP(V_2) / (p_{O_2} * (1 - p_{O_1})) + OP(\emptyset) / ((1 - p_{O_1}) * (1 - p_{O_2})) \end{aligned}$$

Les deux tableaux suivants récapitulent les différents opérateurs d'agrégation  $OP(V_1, V_2)$  (moyenne, somme, nombre, minimum, maximum, écart-type, variance) respectivement pour le cas possibiliste et le cas probabiliste :

Tableau 15 : valeurs possibles des opérateurs d'agrégation classiques pour le cas possibiliste

Opérateur OP	Valeur de l'opérateur d'agrégation $OP(V_1, V_2)$
AVG	$\frac{V_1 + V_2}{2} / (\pi_{O_1} \min \pi_{O_2}) + V_1 / \pi_{O_1} + V_2 / \pi_{O_2} + \emptyset / ((1 - N_{O_1}) \min(1 - N_{O_2}))$

SUM	$(V_1 + V_2)/(\pi_{o_1} \min \pi_{o_2}) + V_1/\pi_{o_1} + V_2/\pi_{o_2} + \emptyset/((1 - N_{o_1}) \min(1 - N_{o_2}))$
COUNT	$2/(\pi_{o_1} \min \pi_{o_2}) + 1/\max(\pi_{o_1}, \pi_{o_2}) + 1/0/((1 - N_{o_1}) \min(1 - N_{o_2}))$
MIN	$\min(V_1, V_2)/(\pi_{o_1} \min \pi_{o_2}) + V_1/\pi_{o_1} + V_2/\pi_{o_2} + \emptyset/((1 - N_{o_1}) \min(1 - N_{o_2}))$
MAX	$\max(V_1, V_2)/(\pi_{o_1} \min \pi_{o_2}) + V_1/\pi_{o_1} + V_2/\pi_{o_2} + \emptyset/((1 - N_{o_1}) \min(1 - N_{o_2}))$
STDDEV	$\text{STDDEV}(V_1, V_2)/(\pi_{o_1} \min \pi_{o_2}) + \text{STDDEV}(V_1)/\pi_{o_1} + \text{STDDEV}(V_2)/\pi_{o_2} + \emptyset/((1 - N_{o_1}) \min(1 - N_{o_2}))$
VAR	$\text{VAR}(V_1, V_2)/(\pi_{o_1} \min \pi_{o_2}) + \text{VAR}(V_1)/\pi_{o_1} + \text{VAR}(V_2)/\pi_{o_2} + \emptyset/((1 - N_{o_1}) \min(1 - N_{o_2}))$

Tableau 16 : valeurs possibles des opérateurs d'agrégation classiques pour le cas probabiliste

Opérateur OP	Valeur OP(V <sub>1</sub> , V <sub>2</sub> )
AVG	$\frac{V_1 + V_2}{2}/(p_{o_1} * p_{o_2}) + V_1/(p_{o_1} * (1 - p_{o_2})) + V_2/(p_{o_2} * (1 - p_{o_1})) + \emptyset/((1 - p_{o_1}) * (1 - p_{o_2}))$
SUM	$(V_1 + V_2)/(p_{o_1} * p_{o_2}) + V_1/(p_{o_1} * (1 - p_{o_2})) + V_2/(p_{o_2} * (1 - p_{o_1})) + \emptyset/((1 - p_{o_1}) * (1 - p_{o_2}))$
COUNT	$2/(p_{o_1} * p_{o_2}) + 1/(p_{o_1} * (1 - p_{o_2}) + p_{o_2} * (1 - p_{o_1})) + 0/((1 - p_{o_1}) * 1 - p_{o_2})$
MIN	$\min(V_1, V_2)/(p_{o_1} * p_{o_2}) + V_1/(p_{o_1} * (1 - p_{o_2})) + V_2/(p_{o_2} * (1 - p_{o_1})) + \emptyset/((1 - p_{o_1}) * (1 - p_{o_2}))$

MAX	$\text{MAX}(V_1, V_2)/(p_{o_1} * p_{o_2}) + V_1/(p_{o_1} * (1 - p_{o_2})) + V_2/(p_{o_2} * (1 - p_{o_1})) + \emptyset / ((1 - p_{o_1}) * (1 - p_{o_2}))$
STDDEV	$\text{STDDEV}(V_1, V_2)/(p_{o_1} * p_{o_2}) + \text{STDDEV}(V_1)/(p_{o_1} * (1 - p_{o_2})) + \text{STDDEV}(V_2)/(p_{o_2} * (1 - p_{o_1})) + \emptyset / ((1 - p_{o_1}) * (1 - p_{o_2}))$
VAR	$\text{VAR}(V_1, V_2)/(p_{o_1} * p_{o_2}) + \text{VAR}(V_1)/(p_{o_1} * (1 - p_{o_2})) + \text{VAR}(V_2)/(p_{o_2} * (1 - p_{o_1})) + \emptyset / (1 - p_{o_1} - p_{o_2})$

Remarque : pour des raisons de clarté et de compréhension des tableaux ci-dessus, nous avons volontairement omis les vecteurs  $X^{I,i}, X_{J,j}$ .

Remarque : les valeurs  $V_1$  et  $V_2$  étant elles-mêmes des valeurs de  $D_{IC}(G)$ , elles sont elles-mêmes des composantes incertaines que l'on devra agréger entre elles, et sur lesquelles on pourra réaliser des opérations algébriques. Nous verrons dans le chapitre suivant que de telles opérations nécessitent des interprétations particulières de combinaisons entre mesures de probabilité et mesures de possibilité.

## 5.6 Conclusion : un formalisme unifié de gestion de connaissances imparfaites

Nous avons défini un formalisme unifié pour la représentation et la manipulation de quantités imparfaites, qui peuvent prendre un nombre fini de valeurs possibles et probables. Si ces quantités sont des valeurs de l'ensemble  $\mathbb{R}$ , nous obtenons une véritable structure algébrique afin d'exploiter l'enchaînement des calculs sur de telles quantités avec des propriétés similaires à  $\mathbb{R}$  que ne possèdent pas les approches classiques basées sur les quantités floues proposées dans la littérature. Les avantages pratiques sont de pouvoir construire rigoureusement une situation avec des dépendances entre imprécisions avec des disjonctions d'expressions. Par ailleurs, les problèmes de fausses extensions de l'univers des possibles sont évités, de même que les termes croisés, lors de calculs sur des quantités mixtes, sont évalués à la fin des calculs, de telle sorte que les données sont agrégées indépendamment de la méthode spécifique de combinaison possibilités/probabilités choisie et que les calculs demeurent rigoureux.

Dans le prochain chapitre, nous proposerons un modèle de décision en contexte incertain. La combinaison des distributions de probabilité avec les distributions de possibilité (somme ou produit) nécessite dans certains cas qu'il convient de définir des transformations particulières de l'une ou autre de ces distributions pour ensuite les agréger en un type unique de distribution. Il existe de nombreuses manières d'interpréter de telles combinaisons via, par exemple, des transformations possibilités/probabilités (ou vice-versa), des mesures de Dempster-Shafer, etc. Nous présenterons enfin des opérateurs d'agrégation génériques pour évaluer plusieurs scénarios possibles en vue de permettre à l'expert de prendre la meilleure décision.

## 6 Chapitre 6 : un modèle de décision en contexte incertain

Dans le chapitre précédent, nous avons proposé un formalisme unifié de gestion de l'incertain. Ce formalisme repose sur l'accumulation de données incertaines qui peuvent ensuite être interprétées par l'expert du domaine en vue d'une prise de décision, par exemple en comparant divers scénarios possibles. L'interprétation particulière de combinaisons de probabilités et de possibilités concerne plus particulièrement le cas de combinaisons de valeurs « mixtes » probabilistes et possibilistes, qui peuvent également être imbriquées. En effet, les termes croisés tels que  $a \times b = \sum_{i=1}^{n_1} \hat{a}_i / \alpha_i . X_{I,i} \times \sum_{j=1}^{n_2} \hat{b}_j / \beta_j . X^{J,j}$  ou  $a = [\text{« bonne »} / 1.X_{I,1} + \text{« très bonne »} / 0.3.X_{I,2}].0.6.X^{J,1} + [\text{« médiocre »} / 1.X_{K,1} + \text{« mauvaise »} / 0.3.X_{K,2}].0.4.X^{J,2}$  n'ont pas été évalués. Ils nécessitent des interprétations de combinaisons de probabilités et de possibilités, plus particulièrement dans le cas de différents types d'imperfections imbriquées. En pratique, ces interprétations doivent être évaluées à la fin des calculs, car ces derniers font perdre les informations concernant les dépendances entre sources d'incertitude et, de plus, ne sont pas toujours bijectives. Dans ce cas, les calculs ne seraient plus rigoureux.

Sur la base du formalisme unifié pour modéliser l'information imparfaite que nous venons de proposer, nous proposons dans ce chapitre un modèle de décision en contexte incertain basé sur les approches présentées ci-après. Tout d'abord, nous proposons différentes approches d'interprétations particulières de combinaisons de valeurs probables et possibles : la théorie de Dempster-Shafer, et les transformations des distributions de probabilité en distributions de possibilité (et vice-versa). Nous proposons ensuite différentes approches de combinaisons des variations sur les valeurs (variateurs dans le cas continu). Par hypothèse, l'incertitude sur l'existence d'une valeur (cas discret représentant les états possibles/probables du monde) est indépendante de la variation autour d'une telle valeur particulière, qui est un variateur continu sous forme d'intervalle, de distribution de probabilité, de possibilité. Enfin, nous proposons un ensemble d'opérateurs génériques d'agrégation pour évaluer et synthétiser les incertitudes

et les différents scénarios possibles en vue de les comparer grâce à une fonction d'utilité, afin que l'expert prenne une décision.

## 6.1 Des interprétations particulières de combinaisons de valeurs entachées d'incertitudes

Dans ce paragraphe, nous formaliserons l'interprétation (ou évaluation) des combinaisons probabilité/possibilité puis nous proposerons deux interprétations particulières de ces combinaisons : l'une basée sur l'approche Dempster-Shafer et la seconde basée sur la transformation normalisation à une en distribution de possibilité. Pour évaluer les expressions obtenues lors des calculs, nous définissons une application que nous noterons  $\varphi$  qui correspond à de telles évaluations :

*Définition : soit  $(a, b) \in (D_{\pi C}(\mathbb{R}) \times D_{pC}(\mathbb{R}))$ ,  $\varphi$  est une application linéaire de  $\mathbb{R} \times (D_{\pi C}(\mathbb{R}) \times D_{pC}(\mathbb{R})) \rightarrow D_{IC}(\mathbb{R})$ , qui associe une distribution incertaine à une interprétation particulière :*

$$\varphi(a \times b) = \varphi \left( \sum_{i=1}^{n_1} \hat{\alpha}_i / \alpha_i \cdot X_{I,i} \times \sum_{j=1}^{n_2} \hat{\beta}_j / \beta_j \cdot X^{J,j} \right)$$

*$\varphi$  est une interprétation particulière de valeurs possibles et probables. Pour chaque combinaison et calcul de quantités de  $D_I(G)$ , nous définissons une matrice des  $\varphi(\alpha_i, \beta_j)$  mise à jour au fur et à mesure des calculs sur chacun des axes  $X_{I,1}, X_{I,2}, \dots, X_{I,n_1}$  d'une part et des axes  $X^{J,1}, X^{J,2}, \dots, X^{J,n_2}$  d'autre part.*

### 6.1.1 Une première interprétation particulière des incertitudes sur les valeurs : l'approche Dempster-Shafer, combinaison conjonctive de valeurs probabilistes et possibilistes

Dans l'état de l'art, nous avons vu que les distributions de probabilité et de possibilités sont des cas particuliers d'une seule et même sur-structure qui est la théorie de l'évidence (ou théorie de Dempster-Shafer). A cet effet, ces deux théorèmes énoncent (1) qu'une distribution de masse (croyance) est une probabilité si et seulement si les éléments focaux sont des singletons et (2) que les éléments focaux sont totalement ordonnés par inclusion si et seulement si Bel et Pl sont respectivement une mesure de nécessité et de possibilité.



Nous avons également vu que l'interprétation des résultats obtenus via les distributions de Dempster-Shafer n'est pas compatible avec les calculs classiques sur des distributions de possibilité : les combinaisons par opérateurs classiques de la logique floue ne sont plus valables et l'ensemble des distributions de possibilité n'est plus stable par combinaison conjonctive des masses de Dempster-Shafer. Cette approche est donc une interprétation particulière de la combinaison probabilités/possibilités. Nous noterons  $\varphi_{DS}$  la fonction d'évaluation « Dempster-Shafer » correspondante.

Nous associons au couple  $(\Pi, P)$  une Mesure de Dempster-Shafer (MDS<sup>37</sup>) définie par les éléments focaux suivants :

$$F_{i,j} = a_i' \times b_j \text{ où } a_i' = \{a_{\chi(1)}, \dots, a_{\chi(i)}\}$$

Où  $\chi$  est une permutation sur  $\{1, 2, \dots, n_1\}$  telle que la suite  $\pi(a_{\chi(1)}), \dots, \pi(a_{\chi(n_1)})$  est décroissante (mais non nécessairement strictement décroissante) avec les masses focales associées :

$$m_{i,j} = m(F_{i,j}) = q_i * p_j$$

Où  $q_i$  et  $p_j$  sont définis par :

- $q_i = \pi(a_{\chi(i)}) - \pi(a_{\chi(i+1)}) = \alpha_{\chi(i)} - \alpha_{\chi(i+1)}$  avec  $0 < i < n_1$ ,  $\alpha_{\chi(1)} = 1$  (la distribution de possibilité est normalisée) et  $q_{n_1} = \alpha_{\chi(n_1)}$ .  $q_i$  est positif ou nul car la suite est décroissante.
- $p_j = \beta_j$ .

La permutation  $\chi$  n'est pas nécessairement unique (dans le cas où plusieurs valeurs sont équipossibles). Nous définissons en fait une famille de MDS dont on peut montrer immédiatement qu'elles conduisent aux mêmes valeurs de plausibilité et crédibilité pour tout sous-ensemble. Il est donc possible de considérer que la MDS définie est l'une de ces mesures équivalentes, représentante de la classe d'équivalence « donne les mêmes valeurs ». La MDS ainsi définie constitue une classe particulière de MDS. Elle sera dite Mesure Produit de Dempster-Shafer (MPDS) de  $\pi$  et de  $p$ , et notée  $m = \pi.p$ .

---

<sup>37</sup> Dans la suite de ce document nous écrirons indifféremment Mesure de Dempster-Shafer ou MDS. Il s'agit en fait de la somme orthogonale de  $\Pi$  et  $P$  considérés comme des MDS

Réciproquement, si une MDS  $m$  est caractérisée par une famille d'ensembles focaux  $F_{i,j}$  avec des masses focales associées  $m_{i,j}$ , et qu'il existe deux ensembles  $\{q_1, \dots, q_{n_1}\}$  et  $\{p_1, \dots, p_{n_2}\}$  vérifiant la relation  $\forall i,j, m_{i,j} = m(F_{i,j}) = q_i * p_j$ , alors  $m$  est une MPDS.

Il existe alors des mesures  $\pi$  et  $p$  uniques telles que  $m = \pi.p$ , qui sera dite décomposition canonique de la D1-distribution  $M$ . La définition de  $m = \pi.p$  correspond à l'hypothèse d'une indépendance entre les connaissances exprimées par  $\pi$  et celles exprimées par  $p$ . Cette hypothèse se formule ainsi :

Définition : soit une distribution de possibilité  $\pi$ , et une distribution de probabilité  $p$ .  $\pi$  et  $p$  seront dites indépendantes si et seulement si :

$$\forall A \neq \emptyset, \forall B \neq \emptyset, A \cap B \neq \emptyset \text{ et } Pl(A \cap B) = \pi(A).p(B) \text{ (plausibilité)}$$

En effet,  $\pi(A)$  définit la borne supérieure d'un intervalle de probabilités. C'est équivalent au fait de dire que c'est vrai pour les éléments focaux de  $\pi$ . Il est immédiat que cette hypothèse conduit à calculer les nouvelles masses associées à  $m$  comme des produits de  $\pi.p$ , et donc à la définition donnée ci-dessus du produit.

Soit  $(a, b) \in (D_{\pi C}(\mathbb{R}) \times D_{pC}(\mathbb{R}))$ . Les calculs de  $\varphi_{DS}$  donnent les résultats suivants :

$$\varphi_{DS}(a \times b) = \varphi_{DS} \left( \sum_{i=1}^{n_1} \hat{a}_i / \alpha_i \cdot X_{I,i} \times \sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{J,j} \right)$$

$$\varphi_{DS}(a \times b) = \left( \sum_{i=1}^{n_1} \{ \hat{a}_{\chi(1)}, \dots, \hat{a}_{\chi(i)} \} / (\alpha_{\chi(i)} - \alpha_{\chi(i+1)}) \cdot X_{I,i} \right) \times \left( \sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{J,j} \right)$$

$$\varphi_{DS}(a \times b) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\{ \hat{a}_{\chi(1)}, \dots, \hat{a}_{\chi(i)} \} / (\alpha_{\chi(i)} - \alpha_{\chi(i+1)}) \cdot X_{I,i}) \times (\hat{b}_j / \beta_j \cdot X^{J,j})$$

$$\varphi_{DS}(a \times b) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\{ \hat{a}_{\chi(1)}, \dots, \hat{a}_{\chi(i)} \} / (\alpha_{\chi(i)} - \alpha_{\chi(i+1)}) \cdot X_{I,i}) \times (\hat{b}_j / \beta_j \cdot X^{J,j})$$

$$\varphi_{DS}(a \times b) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\{ \hat{a}_{\chi(1)}, \dots, \hat{a}_{\chi(i)} \} \times \hat{b}_j) / (\alpha_{\chi(i)} - \alpha_{\chi(i+1)}) * \beta_j \cdot X_{I,i} \cdot X^{J,j}$$

Remarque : il est intéressant de remarquer que  $m$  « tend » (au sens d'une métrique de moindres carrés ou de Sup) vers une distribution de possibilité ou une distribution de

probabilité si une valeur devient prépondérante vis-à-vis des autres dans l'une de ces mesures (i.e. de plus en plus « certaine »), ce qui peut être exprimé par :

$$|\pi(a_{i_0})|_{i \neq i_0} < \varepsilon \text{ (probabilité prépondérante) ou bien :}$$

$$|p(b_{i_0})|_{i \neq i_0} < \varepsilon \text{ (possibilité prépondérante).}$$

On définit alors deux sous-classes particulières de D1-distributions, les purement probables  $(\pi_0, p)$  et les purement possibles  $(\pi, p_0)$  où  $\pi_0$  (respectivement  $p_0$ ) sont les distributions de possibilité (respectivement probabilités) avec un élément totalement nécessaire (respectivement totalement probable) et les autres impossibles (respectivement totalement improbables).

Ainsi, l'opérateur  $\times$  sur les  $(\alpha_i, \beta_j)$  correspond en fait à l'opérateur de combinaison conjonctive de Dempster-Shafer.

### Exemples

(1) Soient  $a$  et  $b$  respectivement une valeur purement probabiliste et purement possibiliste.

- $a = 2$  (probabilité 0.7) ou 3 (probabilité 0.3)
- $b = 1$  (possibilité 1) ou 2 (possibilité 0.5)

$a$  et  $b$  s'expriment de la manière suivante :

- $a = 2 / 0.7.X^{J,1} + 3 / 0.3.X^{J,2}$
- $b = 1 / 1.X_{I,1} + 2 / 0.5.X_{I,2}$

Le cas général avant évaluation s'exprime de la manière suivante :

- $a \times b = (2 / 0.7.X^{J,1} + 3 / 0.3.X^{J,2}) \times (1 / 1.X_{I,1} + 2 / 0.5.X_{I,2})$

L'évaluation  $\varphi_{DS}^{38}$  via la transformation en MDS donne :

- $\varphi_{DS}(a \times b) = (2 / 0.7.X^{J,1} + 3 / 0.3.X^{J,2}) \oplus (1 / 1.X_{I,1} + 2 / 0.5.X_{I,2}) = \varphi_{DS}((2 / 0.7.X^{J,1} + 3 / 0.3.X^{J,2}) \times (1 / 1.X_{I,1} + 2 / 0.5.X_{I,2}))$

On convertit la distribution possibiliste  $b$  en une distribution de masses, ce qui implique :  
 $m(\{1\}) = 0.5 ; m(\{1,2\}) = 0.5$

---

<sup>38</sup> On pourra également noter  $a \oplus b$  l'évaluation  $\varphi_{DS}(a \times b)$

- $\varphi_{DS}(a \times b) = (2 / 0.7.X^{J,1} + 3 / 0.3.X^{J,2}) \oplus (1 / 0.5.X_{I,1} + \{1, 2\} / 0.5.X_{I,2})$
- $\varphi_{DS}(a \times b) = (2 / 0.7.X^{J,1}) \times (1 / 0.5.X_{I,1}) + (2 / 0.7.X^{J,1}) \times (\{1, 2\} / 0.5.X_{I,2}) + (3 / 0.3.X^{J,2}) \times (1 / 0.5.X_{I,1}) + (3 / 0.3.X^{J,2}) \times (\{1, 2\} / 0.5.X_{I,2})$
- $\varphi_{DS}(a \times b) = (2 \times 1) / (0.7.X^{J,1}).(0.5.X_{I,1}) + (2 \times \{1, 2\}) / (0.7.X^{J,1}).(0.5.X_{I,2}) + (3 \times 1) / (0.3.X^{J,2}).(0.5.X_{I,1}) + (3 \times \{1, 2\}) / (0.3.X^{J,2}).(0.5.X_{I,2})$

(2) Dans le cas où  $\times$  est l'opérateur algébrique addition (noté «  $+_{DS}$  »), on obtient :

- $a +_{DS} b = 3/0.35.X_{I,1}.X^{J,1} + \{3,4\}/0.35.X_{I,2}.X^{J,1} + 4/0.15.X_{I,1}.X^{J,2} + \{4,5\}/0.15.X_{I,2}.X^{J,2}$

$a +_{DS} b$  peut prendre trois valeurs possibles : 3, 4 ou 5 avec pour plausibilités respectives : 0.7, 0.65 et 0.15.

(3) Dans le cas où  $\oplus$  est l'opérateur algébrique multiplication (noté «  $*_{DS}$  »), on obtient :

- $a *_{DS} b = 2/0.35.X_{I,1}.X^{J,1} + \{2,4\}/0.35.X_{I,2}.X^{J,1} + 3/0.15.X_{I,1}.X^{J,2} + \{3,6\}/0.15.X_{I,2}.X^{J,2}$

$a *_{DS} b$  peut prendre trois valeurs possibles : 2, 3, 4 ou 6 avec pour plausibilités respectives : 0.7, 0.15, 0.35 et 0.15.

### 6.1.2 Une seconde interprétation particulière : transformation normalisation à un en distribution de possibilité

La transformation normalisation à un en distribution de possibilité est une transformation très simple qui consiste à normaliser la distribution de probabilité par la probabilité la plus élevée [Lasserre, 1999].

Soit  $b = (\hat{b}_i/\beta_i)_{i=1}^n$  une distribution de probabilité avec  $\hat{b}_i$  des éléments de  $G$ , chacun étant affecté d'une probabilité  $\beta_i$ . On définit donc une fonction  $\varphi_{NP}$  de  $\mathbb{R}^n \rightarrow \mathbb{R}$  transformant les coefficients probabilistes en coefficients possibilistes par normalisation à un, définie par :

$$\begin{cases} \varphi_{NP}(\beta_i) = \min \left( \frac{\beta_i}{\max(\beta_i)_{i=1}^n} \right) & \text{dans le cas de coefficients probabilistes} \\ \text{Application identité sinon} \end{cases}$$

$\varphi_{NP}(\beta_i)$  est la possibilité que  $b$  prenne la valeur  $\hat{b}_i$ .

Soit  $(a, b) \in D_{\pi C}(\mathbb{R}) \times D_{pC}(\mathbb{R})$ . Les calculs de  $\varphi_{NP}$  donnent les résultats suivants :

$$\varphi_{NP}(a \times b) = \varphi_{NP} \left( \sum_{i=1}^{n_1} \hat{a}_i / \alpha_i . X_{I,i} \times \sum_{j=1}^{n_2} \hat{b}_j / \beta_j . X^{J,j} \right)$$

$$\varphi_{NP}(a \times b) = \varphi_{NP}\left(\sum_{i=1}^{n_1} \hat{a}_i / \alpha_i \cdot X_{I,i}\right) \times \varphi_{NP}\left(\sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{J,j}\right)$$

$$\varphi_{NP}(a \times b) = \left(\sum_{i=1}^{n_1} \hat{a}_i / \alpha_i \cdot X_{I,i}\right) \times \left(\sum_{j=1}^{n_2} \hat{b}_j / \min\left(\frac{\beta_j}{\max(\beta_j)_{j=1}^{n_2}}\right) \cdot X^{J,j}\right)$$

### Exemples

(1) Soient  $a$  et  $b$  respectivement une valeur purement possibiliste et purement probabiliste.

- $a = 1$  (possibilité 1) ou 2 (possibilité 0.5)
- $b = 2$  (probabilité 0.7) ou 3 (probabilité 0.3)

$a$  et  $b$  s'expriment de la manière suivante :

- $a = 1 / 1.X_{I,1} + 2 / 0.5.X_{I,2}$
- $b = 2 / 0.7.X^{J,1} + 3 / 0.3.X^{J,2}$

Ensuite on additionne ou multiplie avec les opérateurs classiques de la logique floue (principe d'extension).

- $a \times b = (1 / 1.X_{I,1} + 2 / 0.5.X_{I,2}) \times (2 / 0.7.X^{J,1} + 3 / 0.3.X^{J,2})$

On applique la transformation normalisation à 1 sur  $a$  :

- $a \times b = (1 / 1.X_{I,1} + 2 / 0.5.X_{I,2}) \times (2 / 1.X^{J,1} + 3 / \frac{3}{7}.X^{J,2})$

On applique ensuite les opérateurs classiques de combinaisons de possibilités :

- $a \times b = 1 \times 2 / 1 + 2 \times 2 / 0.5 + 1 \times 3 / \frac{3}{7} + 2 \times 3 / \frac{3}{7}$

D'où :

- $a + b = 3 / 1 + 4 / 0.5 + 5 / \frac{3}{7}$  (cas de la somme)
- $a * b = 2 / 1 + 3 / \frac{3}{7} + 4 / 0.5 + 6 / \frac{3}{7}$  (cas du produit)

### 6.1.3 Une première interprétation particulière des variations de valeurs : approche par transformation de distributions de possibilité en famille de probabilités

Nous avons proposé un formalisme pour représenter de manière rigoureuse les états des incertitudes et variations sur les valeurs pouvant hypothétiquement être prises par les distributions de  $D_{IC}(\mathbb{R})$ .

Dans cette partie nous nous intéressons à la partie (2) de l'incertitude, i.e. à la variation autour d'une valeur particulière qui peut être une variation continue sous forme d'intervalle, de distribution de probabilité ou encore de distribution de possibilité. Dans cette approche, nous interprétons les types d'éléments suivants :

Soit  $(a, b) \in (D_{\pi C}(\mathbb{R}) \times D_{pC}(\mathbb{R}))$ . Les calculs de  $\varphi_{NP}$  donnent les résultats suivants :

$$\varphi_{NP}(a \times b) = \varphi_{NP} \left( \sum_{i=1}^{n_1} \hat{a}_i / \alpha_i \cdot X_{I,i} \times \sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{I,j} \right)$$

$$\varphi_{NP}(a \times b) = \varphi_{NP} \left( \sum_{i=1}^{n_1} \hat{a}_i / \alpha_i \cdot X_{I,i} \right) \times \varphi_{NP} \left( \sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{I,j} \right)$$

Nous proposerons deux interprétations particulières de ces combinaisons : l'une basée sur l'approche par transformation de distributions de possibilité en famille de probabilités et la seconde basée sur l'approche par transformation de distributions de probabilité en distributions de possibilité.

La distribution de possibilité (variateur possibiliste)  $a_i \cdot T(v_i, \lambda_i)$  engendre une famille de probabilités  $F_p$  encadrées (1) par une distribution de probabilité basse qui correspond à une mesure de nécessité et (2) par une distribution de probabilité haute qui correspond à une mesure de possibilité dont les fonctions de répartition sont respectivement notées  $F_*$  et  $F^*$ . Le schéma suivant illustre l'ensemble des fonctions de répartition de cette famille de probabilités avec les notations précédentes :

### Fonctions de répartitions de la famille de probabilités

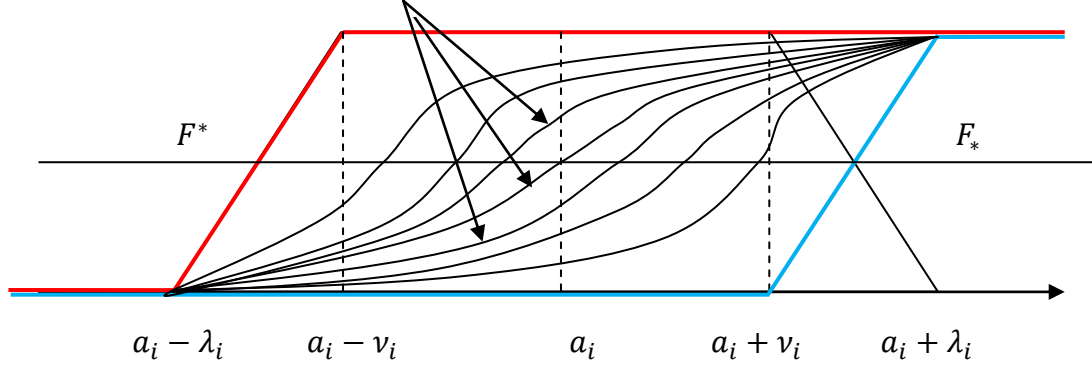


Figure 22 : famille de probabilités induite par une distribution de possibilité

Les deux distributions de probabilité haute et basse encadrant  $F_p$  sont deux lois uniformes respectivement sur les intervalles  $[a_i - \lambda_i, a_i - \nu_i]$  et  $[a_i + \nu_i, a_i + \lambda_i]$ . Elles sont pour espérances respectives  $a_i - \frac{\nu_i + \lambda_i}{2}$  et  $a_i + \frac{\nu_i + \lambda_i}{2}$ , et pour variances (identiques)  $\frac{(\lambda_i - \nu_i)^2}{12}$ . Enfin, nous les noterons respectivement  $a_i.U^*$  et  $a_i.U_*$ . L'espérance de la famille de probabilités est donc comprise dans l'intervalle  $\left[a_i - \frac{\nu_i + \lambda_i}{2}, a_i + \frac{\nu_i + \lambda_i}{2}\right]$ .

On pose  $\theta_1 = a_i - \lambda_i$  et  $\theta_2 = a_i + \lambda_i$ . La variance pour une loi continue sur  $[\theta_1, \theta_2]$  a pour borne supérieure  $\frac{(\theta_2 - \theta_1)^2}{4}$ . Cette dernière correspond à une VA discrète  $X$  avec  $P(X = \theta_1) = 1/2$  et  $P(X = \theta_2) = 1/2$ . La borne inférieure est 0 (VA discrète  $X$  avec  $P(X = \frac{\theta_1 + \theta_2}{2}) = 1$ ).

La variance de la famille de probabilités est donc comprise dans l'intervalle  $[0, \lambda_i^2]$ .

Remarque : toutes les distributions de probabilité  $P$  telles que  $F^* \leq P \leq F_*$  ne vérifient pas la relation  $P \leq \Pi$ . Exemple : soit une valeur floue triangulaire de support  $[1, 3]$  et de mode 2. Soit  $P$  définie par  $P(\{1.5\}) = P(\{2.5\}) = 0.5$ . On a bien  $F^* \leq P \leq F_*$  alors que  $P \notin P(\Pi)$ . En effet,  $P(\{1.5, 2.5\}) = 1 > \Pi(\{1.5, 2.5\}) = 0.5$ . Ainsi la représentation des familles de probabilités par des intervalles flous est davantage informatif que l'encadrement par une paire de fonctions de répartition haute et basse.

$$\varphi_{TPP}(a \times b) = \varphi_{TPP}\left(\sum_{i=1}^{n_1} \hat{a}_i / \alpha_i \cdot X_{I,i}\right)$$

$$\varphi_{TPP}(a \times b) = \sum_{i=1}^{n_1} \varphi_{TPP}(\hat{a}_i) / \alpha_i \cdot X_{I,i}$$

Alors :

$$\varphi_{TPP}(\hat{a}_i) = \varphi_{TPP}(a_i \cdot T(v_{1i}, \lambda_{1i}))$$

$$\varphi_{TPP}(\hat{a}_i) = a_i \cdot \varphi_{TPP}(T(v_{1i}, \lambda_{1i}))$$

$$\varphi_{TPP}(\hat{a}_i) = a_i \cdot \{V / U [-\lambda_i, -v_i] < V < U [v_i, \lambda_i]\}$$

#### 6.1.4 Une seconde interprétation particulière des variations de valeurs : approche par transformation de distributions de probabilité en distributions de possibilité

De nombreuses approches de transformations probabilité-possibilité dans le cas continu ont été proposées dans la littérature : la transformation optimale [Dubois et Prade, 1983], la transformation de Klir [Klir et Parviz, 1992], la nouvelle transformation de Klir [Klir et Wierman, 1998], la transformation normalisation à un ([Palm, 1994], [Bharathi Devi et Sarma, 1985], la transformation de Ree et Jager [Ree et Jager, 1993], etc. Nous avons choisi d'illustrer notre approche avec cette dernière qui est spécifique aux distributions gaussiennes, qui sont alors transformées en distributions de possibilité triangulaire.

La règle de conversion d'une gaussienne  $a_1 \cdot N(\sigma_1^2)$  est la suivante :

$$\varphi_{TPP}(a_1 \cdot N(\sigma_1^2)) = a_1 \cdot T(0, 1.96 \cdot \sigma_1)$$

NB :  $0, 1.96 \cdot \sigma_1$  correspond à l'intervalle de confiance ayant un degré de confiance de 95%.

NB : pour cette transformation, la condition de cohérence (c'est-à-dire que  $P > \Pi$ ) est violée sur les intervalles  $] -\infty, m - 1.96\sigma]$  et  $[m + 1.96\sigma, +\infty[$ .

#### 6.1.5 Cas particuliers : combinaisons du cas continu et du cas discret

Ensuite, deux cas particuliers correspondant à la combinaison d'un variateur continu (sur  $D_{IC}(\mathbb{R})$ ) avec un variateur discret (sur  $D_I(G)$ ) peuvent se présenter :

*Cas particulier 1 : variateur possibiliste discret (i.e.  $v_i = \lambda_i = 0$ ).*



Nous sommes dans le cas d'un variateur possibiliste de type « Dirac », i.e.  $\delta_0 = V(0) = T(0) = 1$  (pas de variation). Dans ce cas,  $a \in D_\pi(\mathbb{R})$  (i.e. cas discret).

Les calculs reviennent à pondérer la distribution de probabilité de loi normale  $b_j.N(\sigma_j^2)$  par un coefficient possibiliste  $\alpha_i$ . Le coefficient possibiliste correspond alors à une valeur discrète  $a_i$  de possibilité  $\alpha_i$ . Les lois de probabilité haute et basse correspondantes sont des lois discrètes de support  $a_i$  de probabilités comprises entre  $N(a_i)$  (nécessité) et  $\alpha_i$  (possibilité). Cela revient donc à pondérer une loi normale  $b_j.N(\sigma_j^2)$  par un coefficient probabiliste  $0 \leq P(a_i) \leq \alpha_i$ .

*Cas particulier 2 : variateur probabiliste discret (i.e.  $\sigma_j = 0$ ).*

Nous sommes dans le cas d'un variateur probabiliste de type « Dirac », i.e.  $\delta_0 = V(0) = N(0) = 1$  (pas de variation). Dans ce cas,  $b \in D_p(\mathbb{R})$  (i.e. cas discret). Les calculs reviennent à pondérer la distribution de possibilité de forme trapézoïdale  $a_i.T(\nu_i, \lambda_i)$  par une probabilité  $\beta_j$ , qui correspond à une valeur discrète  $b_j$ . Un degré de possibilité étant la borne supérieure d'un degré de probabilité, cela revient donc à pondérer une distribution de possibilité de forme trapézoïdale  $a_i.T(\nu_i, \lambda_i)$  par un coefficient possibiliste  $\pi(b_j) \geq \beta_j$ .

Nous venons de proposer différents types d'interprétations particulières pour la combinaison de différents types d'incertitudes. L'intérêt est de pouvoir agréger différentes types d'incertitudes dans un même formalisme : soit une mesure de possibilité (en interprétant les probabilités comme des mesures de possibilités), soit une mesure de probabilité (en interprétant les possibilités comme des mesures de probabilités), soit une mesure de Dempster-Shafer. Dans la seconde partie de ce chapitre, nous proposons un modèle d'agrégation des critères et valeurs en jeu, de sorte que l'expert puisse comparer les différents scénarios dont les utilités potentielles sont synthétisées et prendre une décision.

## 6.2 Des opérateurs génériques d'agrégation

Dans ce paragraphe, nous proposons un ensemble d'opérateurs génériques d'agrégation permettant d'évaluer les incertitudes et l'utilité des différents scénarios possibles. Les distributions et combinaisons de valeurs possibles ou probables sont classiquement agrégées de manière à obtenir une unique valeur évaluant l'utilité de chaque scénario, en vue de prendre des décisions.

## 6.2.1 Distributions de probabilité : cas discret

Soit  $b = \sum_{j=1}^{n_2} b_j / \beta_j \cdot X^{J,j}$ .

Opérations sur les  $b_j$  :

Tableau 17 : opérateurs sur les quantités probabilistes discrètes (1)

Opérateur	Expression	Valeur
Espérance	$E(b_j)$	$b_j$
Probabilité	$proba(b_j)$	$\beta_j$
Probabilité inférieure	$probaInf(b_j)$	
Probabilité supérieure	$probaSup(b_j)$	

Opérations sur  $b$  :

Tableau 18 : opérateurs sur les quantités probabilistes discrètes (2)

Nom de l'opérateur	Notation de l'opérateur	Valeur
Espérance	$E(b)$	$\sum_{j=1}^{n_2} \beta_j \cdot b_j$
Compteur du nombre d'éléments	$count(b)$	$count(b_j / \beta_j \neq 0)_{j=1}^{j=n_2}$
Minimum	$min(b)$	$min(b_j / \beta_j \neq 0)_{j=1}^{j=n_2}$
Maximum	$max(b)$	$max(b_j / \beta_j \neq 0)_{j=1}^{j=n_2}$
Ecart-type	$stddev(b)$	$\sum_{j=1}^{n_2} \beta_j \cdot (b_j - E(b))^2$
Variance	$variance(b)$	$\sqrt{\sum_{j=1}^{n_2} \beta_j \cdot (b_j - E(b))^2}$

## 6.2.2 Distributions de probabilité : cas continu

Soit  $\hat{b} = \sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{J,j}$  avec  $\hat{b}_j = b_j \cdot N(\sigma_j^2)$ .

Opérations sur les  $\hat{b}_j$  :

Tableau 19 : opérateurs sur les quantités probabilistes continues (1)

Opérateur	Expression	Valeur
Espérance	$E(\hat{b}_j)$	$b_j$
Ecart-type	$stddev(\hat{b}_j)$	$\sigma_j$
Variance	$variance(\hat{b}_j)$	$\sigma_j^2$
Probabilité	$proba(c \leq \hat{b}_j \leq d)$	$\int_c^d \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - b_j}{\sigma_j}\right)^2\right) \cdot dx$

Opérations sur  $\hat{b}$  :

Tableau 20 : opérateurs sur les quantités probabilistes continues (2)

Nom de l'opérateur	Notation de l'opérateur	Valeur
Espérance	$E(\hat{b})$	$\sum_{j=1}^{n_2} \beta_j \cdot b_j$
Compteur du nombre d'éléments	$count(\hat{b})$	$count(\hat{b}_j / \beta_j \neq 0)_{j=1}^{j=n_2}$
Minimum	$min(\hat{b})$	$min(\hat{b}_j / \beta_j \neq 0)_{j=1}^{j=n_2}$
Maximum	$max(\hat{b})$	$max(\hat{b}_j / \beta_j \neq 0)_{j=1}^{j=n_2}$
Ecart-type	$stddev(\hat{b})$	$\sqrt{\sum_{j=1}^{n_2} \beta_j^2 \cdot \sigma_j^2 + 2 \sum_{1 \leq i \neq j \leq n_2} \beta_i \cdot \beta_j \cdot cov(\hat{b}_i, \hat{b}_j)}$

Nom de l'opérateur	Notation de l'opérateur	Valeur
Variance	$variance(\hat{b})$	$\sum_{j=1}^{n_2} \beta_j^2 \cdot \sigma_j^2 + 2 \sum_{1 \leq i \neq j \leq n_2} \beta_i \cdot \beta_j \cdot cov(\hat{b}_i, \hat{b}_j)$

Calcul de l'espérance de  $\hat{b}$  :

$$E(\hat{b}) = E\left(\sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{I,j}\right) = E\left(\sum_{j=1}^{n_2} \beta_j \cdot \hat{b}_j\right) = \sum_{j=1}^{n_2} \beta_j \cdot E(\hat{b}_j) = \sum_{j=1}^{n_2} \beta_j \cdot b_j$$

Calcul de la variance de  $\hat{b}$  :

$$V(\hat{b}) = V\left(\sum_{j=1}^{n_2} \hat{b}_j / \beta_j \cdot X^{I,j}\right) = V\left(\sum_{j=1}^{n_2} \beta_j \cdot \hat{b}_j\right)$$

$$V(\hat{b}) = \sum_{j=1}^{n_2} V(\beta_j \cdot \hat{b}_j) + 2 \sum_{1 \leq i \neq j \leq n_2} cov[(\beta_i \cdot \hat{b}_i), (\beta_j \cdot \hat{b}_j)]$$

$$V(\hat{b}) = \sum_{j=1}^{n_2} \beta_j^2 \cdot \sigma_j^2 + 2 \sum_{1 \leq i \neq j \leq n_2} \beta_i \cdot \beta_j \cdot cov(\hat{b}_i, \hat{b}_j)$$

### 6.2.3 Distributions de possibilité : cas discret

Soit  $a = \sum_{i=1}^{n_1} a_i / \alpha_i \cdot X_{I,i}$

Soit  $\chi$  une permutation sur  $\{1, 2, \dots, n_1\}$  telle que la suite  $\pi(a_{\chi(1)}), \dots, \pi(a_{\chi(n_1)})$  est décroissante (mais non nécessairement strictement décroissante). On a alors  $a = \sum_{i=1}^{n_1} a_{\chi(i)} / \alpha_{\chi(i)} \cdot X_{I,\chi(i)}$  avec  $\pi(a_{\chi(1)}) = 1$ .

Rappel : l'interprétation MDS (Mesure de Dempster-Shafer) de  $a$  correspond à la fonction  $\varphi_{DS}$  d'interprétation suivante :

$$\varphi_{DS}(a) = \left( \sum_{i=1}^{n_1} \{a_{\chi(1)}, \dots, a_{\chi(i)}\} / (\alpha_{\chi(i)} - \alpha_{\chi(i+1)}) \cdot X_{I,i} \right)$$

Avec  $\{a_{\chi(1)}, \dots, a_{\chi(i)}\}$  les éléments focaux et  $(\alpha_{\chi(i)} - \alpha_{\chi(i+1)})$  leurs masses respectives.

Opérations sur les  $a_i$  :

Tableau 21 : opérateurs sur les quantités possibilistes discrètes (1)

Nom de l'opérateur	Notation de l'opérateur	Valeur
Probabilité inférieure	$probaInf(a_{\chi(i)})$ $= Bel(a_{\chi(i)})$ $= N(a_{\chi(i)})$	Si $\alpha_{\chi(1)} = 1$ et $\alpha_{\chi(2)} \neq 1$ : $i = 1 \Rightarrow getProbaInf(a_{\chi(1)})$ $= N(a_{\chi(1)})$ $= Bel(a_{\chi(1)})$ $= 1 - \alpha_{\chi(2)}$  $i \in [2, n_1] \Rightarrow getProbaInf(a_{\chi(i)}) = 0$
		Si $\alpha_{\chi(1)} = \alpha_{\chi(2)} = 1$ : $\forall i \in [1, n_1], getProbaInf(a_{\chi(i)}) = 0$
Probabilité supérieure	$probaSup(a_i) = \Pi(a_i)$ $= Pl(a_i)$	$\alpha_i$
Probabilité	$proba(a_{\chi(i)})$ $= betP(a_{\chi(i)})$	$\frac{1}{1 - m_{\cap}(\phi)} \sum_{\omega \in A} \frac{(\alpha_{\chi(i)} - \alpha_{\chi(i+1)})}{ \{a_{\chi(1)}, \dots, a_{\chi(i)}\} }$

Opérations sur  $a$  :

Tableau 22 : opérateurs sur les quantités possibilistes discrètes (2)

Nom de l'opérateur	Notation de l'opérateur	Valeur
Espérance	$E(a)$	$\sum_{i=1}^{n_1} getProba(a_{\chi(i)}) \cdot a_i$
Compteur du nombre d'éléments	$count(a)$	$count(a_i / \alpha_i \neq 0)_{i=1}^{i=n_1}$

Nom de l'opérateur	Notation de l'opérateur	Valeur
Minimum	$\min(a)$	$\min(a_i/\alpha_i \neq 0)_{i=1}^{i=n_1}$
Maximum	$\max(a)$	$\max(a_i/\alpha_i \neq 0)_{i=1}^{i=n_1}$
Ecart-type	$stddev(a)$	$\sum_{j=1}^{n_2} \text{getProba}(a_{\chi(i)}) \cdot (a_{\chi(i)} - E(a))^2$
Variance	$variance(a)$	$\sqrt{\sum_{j=1}^{n_2} \text{getProba}(a_{\chi(i)}) \cdot (a_{\chi(i)} - E(a))^2}$

#### 6.2.4 Distributions de possibilité : cas continu

Soit  $\hat{a} = \sum_{i=1}^{n_1} \hat{a}_i/\alpha_i \cdot X_{I,i}$  avec  $\hat{a}_i = a_i \cdot T(v_i, \lambda_i)$

Opérations sur les  $\hat{a}_i$  :

Tableau 23 : opérateurs sur les quantités possibilistes continues (1)

Opérateur	Expression	Valeur
Espérance	$E(\hat{a}_i)$	$\hat{a}_i$
Valeur centrale	$centralValue$	$a_i$
Noyau	$kernel(\hat{a}_i)$	$[a_i - v_i, a_i + v_i]$
	$kernelMin(\hat{a}_i)$	$a_i - v_i$
	$kernelMax(\hat{a}_i)$	$a_i + v_i$
Support	$support(\hat{a}_i)$	$[a_i - \lambda_i, a_i + \lambda_i]$
	$supportMin(\hat{a}_i)$	$a_i - \lambda_i$
	$supportMax(\hat{a}_i)$	$a_i + \lambda_i$
Probabilité inférieure	$probaInf(\hat{a}_i) = Bel(\hat{a}_i)$ $= N(\hat{a}_i)$	0

Opérateur	Expression	Valeur
Probabilité supérieure	$probaSup(\hat{a}_i) = \Pi(\hat{a}_i)$ $= Pl(\hat{a}_i)$	1

Une interprétation de  $a_i.T(v_i, \lambda_i)$  en famille de probabilités donne les opérateurs d'agrégation supplémentaires suivants :

Tableau 24 : opérateurs sur les quantités possibilistes discrètes (2)

Opérateur	Expression	Valeur
Espérance inférieure	$E_{inf}(\hat{a}_i)$	$a_i - \frac{v_i + \lambda_i}{2}$
Espérance supérieure	$E_{sup}(\hat{a}_i)$	$a_i + \frac{v_i + \lambda_i}{2}$
Variance	$variance(\hat{a}_i)$	$[0, \lambda_i^2[$
Variance inférieure	$varianceInf(\hat{a}_i)$	0
Variance supérieure	$varianceSup(\hat{a}_i)$	$\lambda_i^2$

### 6.3 Conclusion : des interprétations de combinaisons des distributions incertaines, des opérateurs génériques d'agrégation pour la prise de décision en contexte incertain

Nous venons de proposer différentes interprétations particulières de combinaisons de valeurs et de variateurs incertains sur la base du formalisme unifié que nous avons précédemment proposé pour modéliser l'information imparfaite. Enfin, nous avons proposé un ensemble d'opérateurs génériques d'agrégation pour évaluer et synthétiser les incertitudes et les différents scénarios possibles pour la prise de décision.

Bien que d'autres interprétations particulières et opérateurs soient disponibles dans la littérature, nous avons choisi celles qui nous semblaient les plus représentatives. L'avantage de notre approche est que le formalisme unifié proposé est compatible avec ces autres interprétations qu'il convient d'implémenter en fonction des besoins.

Dans la prochaine partie (« validation »), nous validerons nos contributions par la mise en œuvre de deux scénarios d'aide à la décision portant sur des exemples liés à l'agriculture, puis nous proposerons une opérationnalisation avec une architecture orientée service que nous réaliserons sur la base des concepts que nous avons précédemment définis.



## Quatrième partie : validation

## 7 Chapitre 7 : évaluation : des scénarios d'aide à la décision sur la période de semis dans une exploitation agricole

Dans ce chapitre, nous appliquons et évaluons les propositions présentées dans la partie précédente. Pour cela, nous mettons en œuvre dans un premier temps deux études de cas qui traitent différents problèmes d'aide à la décision sur la période de semis dans une exploitation agricole. Dans un second temps, nous opérationnalisons nos contributions avec une architecture orientée service basée sur les concepts précédemment définis, et plus particulièrement en rapport avec les différents types d'imperfections qui peuvent entacher des données.

### 7.1 Étude de cas 1 : de l'utilité de distinguer probabilités et possibilités

Cette étude de cas présente l'intérêt de traiter les différents types d'incertitudes (i.e. les possibilités et les probabilités) de manière différente.

#### *Présentation du cas*

Un exploitant agricole possède un semoir. Le semoir sème une certaine quantité de semences par unité de surface et à une certaine profondeur. Pour semer une parcelle avec un type de culture donné, l'exploitant agricole a deux choix : le semis précoce (scénario A) ou le semis tardif (scénario B). Le fabricant de semences indique les quantités à semer pour chacun des scénarios : elles sont respectivement de 160kg/ha et 140kg/ha. Toutes les semences sont supposées identiques. La densité de semis est supposée proportionnelle à la productivité par hectare. Ainsi, on suppose l'existence d'une fonction d'utilité égale à 100 fois la masse de graines semées par hectare (densité de semis) pour chacun des scénarios.

Le semis précoce (scénario A) nécessite de semer davantage en surface (recherche du réchauffement du lit de semences) tandis que le semis tardif nécessite de semer en profondeur (afin que les racines séminales trouvent de l'humidité lors de la germination). La profondeur

de semis a cependant une incidence sur la productivité. La productivité dépend donc à la fois de la profondeur et de la densité du semis. En termes mathématiques, on a donc une fonction  $f$  qui à la profondeur du semis « Prof » et la densité de semis « Ds » associe la productivité Prod :

$$Prod = f(Prof, Ds)$$

- Dans le scénario A, semer en profondeur moyenne procure un gain de 10% ; semer en profondeur élevée procure un gain de 5%.
- Dans le scénario B, semer en profondeur faible procure un gain de 20% ; semer en profondeur moyenne procure un gain de 10%.
- Dans tous les autres cas, le gain est neutre (0%).

Le premier problème est que la quantité de graines semées par le semoir pour chaque unité de surface est aléatoire malgré un réglage précis. On va pouvoir cependant mesurer la quantité de graines a posteriori car certaines graines ont des nanocapteurs permettant de les géolocaliser en surface. Cependant, les graines ne peuvent pas être géolocalisées en profondeur.

Cela conduit au deuxième problème qui est que la profondeur de mise en terre des semences par le semoir n'est pas réglée de manière précise. On sait simplement que si l'on règle la machine en mode « proche surface » (scénario A), il est tout à fait possible ( $\Pi = 1, N = 0.7$ ) que la profondeur soit faible mais il est aussi possible ( $\Pi = 0.3, N = 0$ ) qu'elle soit moyenne pour toutes les graines semées. Dans le cas du mode « en profondeur » (scénario B), il est tout à fait possible ( $\Pi = 1, N = 0.4$ ) que la profondeur soit élevée mais il est aussi possible ( $\Pi = 0.6, N = 0$ ) qu'elle soit très élevée pour toutes les graines semées.

Voici un tableau récapitulatif des deux scénarios possibles :

Tableau 25 : récapitulatif des deux scénarios possibles

Scénarios \ semences	Qté semée (données fabricant)	Profondeur de semis	Profondeur	Gain associé	Possibilité Nécessité
Scénario A	160 kg/ha	Surface	Faible	10%	$\Pi = 1$ $N = 0.7$
			Moyenne	0%	$\Pi = 0.3$ $N = 0$

Scénario B	140 kg/ha	Profondeur	Élevée	30%	$\Pi = 1$ $N = 0.4$
			Très élevée	10%	$\Pi = 0.6$ $N = 0$

On modélise alors les différents scénarios en fonction de la quantité et de la qualité de mesures dont on dispose.

### 7.1.1 Cas 1 : distributions de possibilité discrètes et valeurs à la fois précises et certaines

Supposons que nous avons à disposition une unique mesure à la fois précise et certaine des quantités de graines semées par unité de surface. On modélise les deux scénarios possibles comme suit :

- Scénario A :  $(faible/1.X_{I,1} + moyenne/0.3.X_{I,2}) \times (160/1.X^{J,1})$
- Scénario B :  $(\acute{e}lev\acute{e}e/1.X_{K,1} + tr\grave{e}s\_ \acute{e}lev\acute{e}e/0.6.X_{K,2}) \times (140/1.X^{L,1})$

#### *Calcul des utilités et prise de décision*

Les utilités des différents croisements entre les cas de profondeur et de quantité à semer sont alors les suivantes :

Tableau 26 : cas 1 : utilités des différents scénarios

Scénarios \ caractéristiques	Profondeur / qté semée	Gain	Utilité
Scénario A	Faible / 160	10%	176
	Moyenne / 160	0%	160
Scénario B	Élevée / 140	30%	182
	Très élevée / 140	10%	154

Les utilités calculées sont précises et certaines. Pour prendre une décision, l'exploitant doit choisir entre les différents critères pour la décision en incertain non mesurable :

- S'il est « prudent », le critère minimax (160) lui suggère de choisir le scénario A.
- S'il est « audacieux », le critère maximax (182) lui suggère de choisir le scénario B.

Pour ces deux critères, les valeurs des probabilités et/ou des possibilités ne sont pas exploitées. En effet, pour les probabilités, il n'y a pas de notion de fréquence puisqu'on a une seule mesure. Le gain de chaque scénario est une distribution de possibilités du type :

- Scénario A : 176/1 + 160/0.3
- Scénario B : 140/1 + 154/0.6

Nous obtenons donc des gains de 176 (possibilité 1, nécessité 0.7) et 160 (possibilité 0.3, nécessité 0) pour le scénario A ; et 140 (possibilité 1, nécessité 0.4) et 154 (possibilité 0.6, nécessité 0) pour le scénario B. Il est également possible d'agréger les gains possibles de chaque scénario en réalisant une interprétation de type MDS<sup>39</sup> puis en pondérant chaque gain par sa probabilité pignistique. On obtient alors « l'utilité espérée MDS » :

- Scénario A :  $BetP(176) = 0.85$  ;  $BetP(160) = 0.15$
- Scénario B :  $BetP(182) = 0.7$  ;  $BetP(154) = 0.3$

Nous obtenons enfin des gains de 173.6 pour chacun des deux scénarios.

### 7.1.2 Cas 2 : distributions de possibilité discrètes et ensembles des valeurs aléatoires

Pour n parcelles (n mesures), l'échantillon des quantités semées devient de plus en plus représentatif lorsque n augmente pour la partie probabiliste (quantité de grains à semer). Nous obtenons un modèle du type suivant :

- Scénario A :

$$(faible/1.X_{I,1} + moyenne/0.3.X_{I,2}) \\ \times \left( \frac{1}{n}/159.X^{L,1} + \frac{1}{n}/162.X^{L,2} + \dots + \frac{1}{n}/164.X^{L,n} \right)$$

(n fois pour que la somme des probabilités soit égale à 1.  $\frac{1}{n}$  varie en fonction du nombre n de mesures.

- Scénario B :

---

<sup>39</sup> Mesure de Dempster-Shafer

$$(\text{élevée}/1.X_{I,1} + \text{très\_élevée}/0.3.X_{I,2}) \\ \times \left( \frac{1}{n}/141.X^{L,1} + \frac{1}{n}/145.X^{L,2} + \dots + \frac{1}{n}/138.X^{L,n} \right)$$

#### Calcul des utilités et prise de décision

Dans ce cas, pour les mesures on peut raisonner sur l'espérance du gain côté probabilités. En effet, d'après la loi des grands nombres, la moyenne empirique tend vers l'espérance réelle et la variance se réduit. Le nombre de mesures étant conséquent, les croisements entre les cas et les utilités correspondantes sont nombreux. En pratique, en vue d'une prise de décision, il faudra dans un premier temps calculer l'espérance des valeurs obtenues lors des mesures.

- Pour le scénario A, on obtiendra par exemple :

$$E \left( \frac{1}{n}/159.X^{L,1} + \frac{1}{n}/162.X^{L,2} + \dots + \frac{1}{n}/164.X^{L,n} \right) = 162$$

- Pour le scénario B, on obtiendra par exemple :

$$E \left( \frac{1}{n}/141.X^{L,1} + \frac{1}{n}/145.X^{L,2} + \dots + \frac{1}{n}/138.X^{L,n} \right) = 139$$

Les utilités des différents croisements entre les cas de profondeur et de quantité à semer sont alors les suivantes :

Tableau 27 : cas 2 : utilités des différents scénarios

Scénarios \ caractéristiques	Profondeur / qté semée espérée	Gain	Utilité
Scénario A	Faible / 162	10%	178.2
	Moyenne / 162	0%	162
Scénario B	Élevée / 139	30%	180.7
	Très élevée / 139	10%	152.9

De la même manière que dans le cas (1), les espérances obtenues seront elles-mêmes des distributions de possibilité sur lesquelles on pourra réaliser un minimax ou un maximax sur les croisements entre les cas de profondeur et de quantité semée « espérée ».

- S'il est « prudent », le critère minimax (162) lui suggère de choisir le scénario A.
- S'il est « audacieux », le critère maximax (180.7) lui suggère de choisir le scénario B.

### Remarque

De la même manière que précédemment, on pourra utiliser une interprétation MDS afin d'évaluer « l'utilité espérée MDS » de chaque scénario.

- Scénario A :  $BetP(178.2) = 0.85$  ;  $BetP(162) = 0.15$
- Scénario B :  $BetP(180.7) = 0.7$  ;  $BetP(152.9) = 0.3$

Nous obtenons alors des gains respectifs de 175.77 et de 172.36 pour les scénarios A et B.

### Discussion

Ce type de problème illustre différents cas où nous traitons les probabilités et les possibilités de manières différentes, plus particulièrement lors de la répétition de phénomènes (par exemple lors de simulations du type  $x_{n+1} = a.x_n + b$ ).

En effet, la principale différence se situe lors de la répétition d'épreuves où intervient la nature aléatoire d'un phénomène, les quantités probabilistes sont modifiées aléatoirement à chaque mesure constituant l'échantillon. Nous remplaçons alors classiquement ces valeurs par leur valeur agrégée telle que l'espérance ou la variance, ce que permet la forme canonique développée qui scinde l'univers des « possibles » de l'univers « probabiliste », de la forme (univers possible) \* (univers probable).

Par contre, lorsque des épreuves possibilistes se répètent ; par exemple si on utilise la même machine sur plusieurs parcelles ou bien si on simule l'utilisation de la même machine sur plusieurs années, on a une agrégation de valeurs possibilistes dont les sources d'incertitudes sont identiques entre elles, i.e. des expressions du type :

$$(faible/1.X_{I,1} + moyenne/0.3.X_{I,2}) \times (faible/1.X_{I,1} + moyenne/0.3.X_{I,2}) \times \dots \\ \times (faible/1.X_{I,1} + moyenne/0.3.X_{I,2}) \text{ (m fois)}$$

D'après le formalisme défini précédemment, on a :  $X_{I,1}.X_{I,1} = X_{I,1}$ ,  $X_{I,2}.X_{I,2} = X_{I,2}$  et  $X_{I,1}.X_{I,2} = 0$ . Le résultat de l'agrégation est donc :

$$(faible \times faible \times \dots \times faible)/1.X_{I,1} \\ + (moyenne \times moyenne \times \dots \times moyenne)/0.3.X_{I,2}$$

Ce qui correspond bien à l'intuition qu'on ne « compte » qu'une seule fois l'incertitude lorsqu'on répète les « épreuves » possibilistes.

### 7.1.3 Cas 3 : distributions de possibilité continues et ensembles des valeurs aléatoires

Plutôt que de décrire qualitativement la profondeur de semis avec une distribution de possibilité discrète, il est plus réaliste de considérer chaque type de profondeur par un nombre flou continu (trapézoïde) dont l'unité de mesure est le cm.

Tableau 28 : cas 3 : lien profondeur de semis / nombre flou

Profondeur	Nombre flou correspondant
Faible	$1. T(0.2, 0.5)$
Moyenne	$2.5. T(0.3, 0.7)$
Elevée	$3.5. T(0.2, 0.3)$
Très élevée	$5. T(0.2, 0.2)$

Nous obtenons un modèle du type suivant :

- Scénario A :

$$(1. T(0.2, 0.5)/1. X_{I,1} + 2.5. T(0.3, 0.7)/0.3. X_{I,2}) \\ \times \left( \frac{1}{n}/159. X^{L,1} + \frac{1}{n}/162. X^{L,2} + \dots + \frac{1}{n}/164. X^{L,n} \right)$$

- Scénario B :

$$(3.5. T(0.2, 0.3)/1. X_{I,1} + 5. T(0.2, 0.2)/0.3. X_{I,2}) \\ \times \left( \frac{1}{n}/141. X^{L,1} + \frac{1}{n}/145. X^{L,2} + \dots + \frac{1}{n}/138. X^{L,n} \right)$$

#### *Exemple d'événement imprévu*

Nous pouvons par exemple exploiter ce cas de figure avec l'exemple suivant : supposons que le scénario A ait été initié et que peu après des oiseaux aient picoré les graines semées, dégradant également le champ. L'agriculteur décide alors de réaliser un semis tardif. Cependant, des imprécisions de  $\pm 0.05 \text{ cm}$  et  $\pm 0.1 \text{ cm}$  auxquelles il faut ajouter algébriquement des biais de  $-0.2 \text{ cm}$  et  $+0.1 \text{ cm}$  sont ajoutées respectivement en cas de profondeur élevée et très élevée.



Côté possibilités, les résultats de calculs sont les suivants :

- Scénario B :  $(3.5.T(0.2,0.3)/1.X_{I,1} + 5.T(0.2,0.2)/0.3.X_{I,2}) \times (-0.2.T(0.05)/1.X_{I,1} + 0.1.T(0.1)/0.3.X_{I,2})$  (avec  $\times$  l'opérateur addition).

Nous obtenons alors la nouvelle distribution des profondeurs de semis :

- Scénario B :  $(3.3.T(0.25,0.35)/1.X_{I,1} + 5.1.T(0.3,0.3)/0.3.X_{I,2})$

Le gain est lui-même un sous-ensemble flou. Nous obtenons donc une distribution de possibilité de sous-ensembles flous.

#### *Calcul des utilités et prise de décision*

Reprenons les scénarios A et B initiaux du cas 3, avec les hypothèses suivantes :

- Les profondeurs respectives de 1, 2.5, 3.5, 5 correspondent respectivement à des gains de 10%, 0%, 30%, 10%.
- L'imprécision de profondeur de semis est proportionnelle à l'imprécision de l'utilité. Ainsi une modification de 1 cm entraîne un gain modifié de 10%.

Concernant les gains, nous obtenons alors les distributions suivantes :

- Scénario A :

$$(10.T(2,5)/1.X_{I,1} + 0.T(2,7)/0.3.X_{I,2}) \\ \times \left( \frac{1}{n}/159.X^{L,1} + \frac{1}{n}/162.X^{L,2} + \dots + \frac{1}{n}/164.X^{L,n} \right)$$

- Scénario B :

$$(30.T(2,3)/1.X_{I,1} + 10.T(2,2)/0.3.X_{I,2}) \\ \times \left( \frac{1}{n}/141.X^{L,1} + \frac{1}{n}/145.X^{L,2} + \dots + \frac{1}{n}/138.X^{L,n} \right)$$

Après calcul de l'espérance côté probabilités, nous obtenons des distributions de possibilité de sous-ensembles flous :

- Scénario A :  $178.2.T(5.3,8.9)/1 + 162.T(3.2,11.3)/0.3$
- Scénario B :  $180.7.T(3.6,5.4)/1 + 152.9.T(3.1,3.1)/0.3$

(Les résultats des calculs ont été arrondis à 0.1 près pour plus de clarté).

### *Discussion*

Les résultats des calculs sont des distributions de possibilités de sous-ensembles flous. Nous pouvons encore exploiter les critères minimax et maximax sur les croisements entre les cas de profondeur et de quantité semée « espérée », en tenant compte de l'ensemble des valeurs possibles.

- S'il est « prudent », le critère minimax ( $162 - 11.3 = 150.7$ ) lui suggère de choisir le scénario A comme lors des cas précédents. Cependant, on remarque le rapprochement des minimax des deux scénarios due à la plus grande imprécision de la valeur  $162.T(3.2, 11.3)$  du scénario A.
- S'il est « audacieux », le critère maximax ( $187.1 = 178.2 + 8.9$ ) lui suggère de choisir le scénario A au lieu du scénario B lors des cas précédents.

### *Calcul de l'espérance*

Avec une interprétation MDS, les éléments focaux sont alors des nombres flous. Nous obtenons alors les valeurs suivantes :

- Scénario A :  $betP(178.2.T(5.3, 8.9)) = 0.85$  ;  $betP(162.T(3.2, 11.3)) = 0.15$
- Scénario B :  $betP(180.7.T(3.6, 5.4)) = 0.7$  ;  $betP(152.9.T(3.1, 3.1)) = 0.3$

Si l'on pondère les éléments focaux par les probabilités pignistiques correspondantes, nous obtenons les résultats suivants :

$$UE(\text{scénario A}) = 151.5.T(4.5, 7.6) + 24.3.T(0.5, 1.7) = 175.8.T(5.0, 9.3)$$

$$UE(\text{scénario B}) = 126.5.T(2.5, 3.8) + 45.9.T(0.9, 0.9) = 172.4.T(3.4, 4.7)$$

Si l'utilité espérée du scénario A est plus élevée, sa plage de valeurs est plus large. En effet, le gain minimax est du côté du scénario B.

### *Remarque*

Les résultats des valeurs centrales sont les mêmes que ceux obtenus dans le cas 2.

### *Synthèse*

Le tableau suivant synthétise les décisions calculées en fonction des critères et des cas.

Tableau 29 : synthèse des décisions calculées en fonction des critères et des cas

Critère	Choix pour le cas 1	Choix pour le cas 2	Choix pour le cas 3
Espérance	A = B	A	A mais plus grande variabilité
Minimax	A	A	A
Maximax	B	B	A

Le calcul de l'espérance revient à transformer la distribution de possibilité en distribution de probabilité. Si un tel calcul permet de synthétiser en une unique valeur l'indicateur utilité, cette approche ne permet pas de prendre en compte la sensibilité du décideur aux gains et aux pertes, i.e. de mettre en évidence toutes les subtilités que pourrait engendrer l'état d'esprit du décideur : il peut-être soit prudent (choix du minimax) soit audacieux (choix maximax).

## 7.2 Étude de cas 2 : cas de probabilités et possibilités imbriquées

Dans ce cas, nous choisissons de traiter un cas avec des possibilités et des probabilités imbriquées. Ce cas illustre et évalue les différents types d'interprétations étudiés précédemment.

### Présentation du cas

Il s'agit du cas précédent sur les semoirs que nous avons modifié pour faire apparaître les imbrications probabilités/possibilités étudiées dans le chapitre précédent. Une information supplémentaire précise que le fabricant de semoirs fournit 70% de machines de type X et 30% de machines de type Y de précisions différentes, impliquant ainsi une dépendance entre une distribution de probabilité et une distribution de possibilité. Sachant que l'exploitant agricole possède plusieurs machines de ce fabricant, il y a donc 70% de chances de semer avec une machine de type X et 30% d'avoir une machine de type Y.

- Scénario A :

$$\left[ (faible/1.X_{I,1} + moyenne/0.3.X_{I,2})/0.7.X^{M,1} + (faible/1.X_{I,1} + moyenne/0.6.X_{I,2})/0.3.X^{M,2} \right] \times (160/1.X^{J,1})$$

- Scénario B :

$$\begin{aligned} & \left[ (\text{élevée}/1.X_{K,1} + \text{très\_élevée}/0.6.X_{K,2})/0.7.X^{M,1} \right. \\ & \quad \left. + (\text{élevée}/1.X_{K,1} + \text{très\_élevée}/0.8.X_{K,2})/0.3.X^{M,2} \right] \times (140/1.X^{L,1}) \end{aligned}$$

En pratique, dans un tel scénario, l'expert pourra être amené à interpréter probabilités et possibilités dans un formalisme commun, ce que nous allons illustrer ci-après avec deux exemples d'interprétations particulières dont nous allons comparer les résultats en termes d'utilités espérée, inférieure et supérieure.

### 7.2.1 Interprétation 1 : mesures de Dempster-Shafer

Dans cette interprétation, il faut convertir les mesures de possibilités en « masses imbriquées ».

$$\begin{aligned} \varphi_{DS}(\text{Scénario } A) &= \left( (\{faible\}/0.7.X_{I,1} + \{faible, moyenne\}/0.3.X_{I,2})/0.7.X^{M,1} \right. \\ & \quad \left. + (\{faible\}/0.4.X_{I,1} + \{faible, moyenne\}/0.6.X_{I,2})/0.3.X^{M,2} \right) \\ & \quad \times (160/1.X^{J,1}) \end{aligned}$$

$$\begin{aligned} \varphi_{DS}(\text{Scénario } B) &= \left( (\{élevée\}/0.4.X_{K,1} + \{élevée, très\_élevée\}/0.6.X_{K,2})/0.7.X^{M,1} \right. \\ & \quad \left. + (\{élevée\}/0.2.X_{K,1} + \{élevée, très\_élevée\}/0.8.X_{K,2})/0.3.X^{M,2} \right) \\ & \quad \times (140/1.X^{L,1}) \end{aligned}$$

Pour la prise de décision, on pourra agréger les quantités disponibles et ainsi évaluer l'utilité espérée globale pour chaque scénario en calculant la moyenne pondérée des utilités de chaque cas probable/possible. Les pondérations pourront se réaliser grâce à trois quantités distinctes :

- La probabilité pignistique<sup>40</sup>
- La crédibilité<sup>41</sup> (que nous appellerons probabilité inférieure)

---

<sup>40</sup> Rappel : quelque soit le singleton  $\omega \in \Omega$ , l'expression de la probabilité pignistique est :

$\text{BetP}(\omega) = \frac{1}{1 - m_\cap(\phi)} \sum_{\omega \in A} \frac{m(A)}{|A|}$  avec  $|A|$  étant le cardinal du sous-ensemble  $A$ .

- La plausibilité<sup>42</sup> (que nous appellerons probabilité supérieure)

Après développement des expressions ci-dessus, nous obtenons les formules développées suivantes :

$$\begin{aligned}
\varphi_{DS}(\text{Scénario } A) &= \left( (\{faible\} \times 160) / 0.49 \cdot X_{I,1} \cdot X^{M,1} \cdot X^{J,1} \right. \\
&\quad + (\{faible, moyenne\} \times 160) / 0.21 \cdot X_{I,2} \cdot X^{M,1} \cdot X^{J,1} \Big) \\
&\quad + \left( (\{faible\} \times 160) / 0.12 \cdot X_{I,1} \cdot X^{M,2} \cdot X^{J,1} \right. \\
&\quad \left. + (\{faible, moyenne\} \times 160) / 0.18 \cdot X_{I,2} \cdot X^{M,2} \cdot X^{J,1} \right)
\end{aligned}$$

$$\varphi_{DS}(\text{Scénario } A) = (\{faible\} \times 160) / 0.61 + (\{faible, moyenne\} \times 160) / 0.39$$

$$\begin{aligned}
\varphi_{DS}(\text{Scénario } B) &= \left( (\{élevée\} \times 140) / 0.28 \cdot X_{K,1} \cdot X^{M,1} \cdot X^{L,1} \right. \\
&\quad + (\{élevée, très\_élevée\} \times 140) / 0.42 \cdot X_{K,2} \cdot X^{M,1} \cdot X^{L,1} \Big) \\
&\quad + \left( \{élevée\} \times 140 / 0.06 \cdot X_{K,1} \cdot X^{M,2} \cdot X^{L,1} + \{élevée, très\_élevée\} \right. \\
&\quad \left. \times 140 / 0.24 \cdot X_{K,2} \cdot X^{M,2} \cdot X^{L,1} \right)
\end{aligned}$$

$$\varphi_{DS}(\text{Scénario } B) = (\{élevée\} \times 140) / 0.34 + (\{élevée, très\_élevée\} \times 140) / 0.66$$

*Calcul des utilités espérées*

Nous en déduisons les probabilités pignistiques des différents événements :

Scénario A :

$$BetP(faible) = 0.49 + 0.105 + 0.12 + 0.09 = 0.805$$

---

<sup>41</sup> Rappel : le degré de croyance Bel de A évalue la quantité totale potentiellement affectée spécifiquement à A.  $\forall A \subseteq \Omega, Bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B)$

<sup>42</sup> Rappel : la mesure de plausibilité Pl de A évalue la quantité maximum d'évidence qui peut être potentiellement affectée à A.  $\forall A \subseteq \Omega, Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$

$$BetP(moyenne) = 0.105 + 0.09 = 0.195$$

Scénario B :

$$BetP(élevée) = 0.28 + 0.21 + 0.06 + 0.12 = 0.67$$

$$BetP(très\_élevée) = 0.21 + 0.12 = 0.33$$

Nous obtenons alors l'utilité espérée de chacun des scénarios :

$$UE(\varphi_{DS}(Scénario A)) = 0.805 * 176 + 0.195 * 160 = 172.88$$

$$UE(\varphi_{DS}(Scénario B)) = 0.67 * 182 + 0.33 * 154 = 172.76$$

*Calcul des probabilités inférieures et supérieures*

A partir de  $\varphi_{DS}(Scénario A)$  et  $Eval_{DS}(Scénario B)$  on calcule :

- $Bel(\{faible\}) = 0.61$  et  $Pl(\{faible\}) = 1$
- $Bel(\{moyenne\}) = 0$  et  $Pl(\{moyenne\}) = 0.39$

On obtient alors:

$$0.61 * 176 + 0 * 160 \leq UE(\varphi_{DS}(Scénario A)) \leq 1 * 176 + 0.39 * 160$$

$$107.36 \leq UE(\varphi_{DS}(Scénario A)) \leq 238.4$$

A partir de  $Eval_{DS}(Scénario B)$  on calcule :

- $Bel(\{élevée\}) = 0.34$  et  $Pl(\{élevée\}) = 1$
- $Bel(\{très\_élevée\}) = 0$  et  $Pl(\{très\_élevée\}) = 0.66$

On obtient alors :

$$0.34 * 182 + 0 * 154 \leq UE(\varphi_{DS}(Scénario B)) \leq 1 * 182 + 0.66 * 154$$

$$61.88 \leq UE(\varphi_{DS}(Scénario B)) \leq 283.64$$

Si l'utilité espérée est quasi identique entre les deux scénarios A et B, les utilités inférieures et supérieures du scénario B induisent un intervalle beaucoup plus large. Le scénario A propose donc des résultats beaucoup moins variables dans le pire ou le meilleur des cas.

Remarque : une interprétation donnant les mêmes résultats est l'interprétation que nous appelons « probabilités imprécises ». Dans cette approche, nous interprétons les possibilités et

nécessités comme encadrant une mesure de probabilité imprécise. Après transformation des mesures de possibilités en mesures de probabilités, on obtient les distributions suivantes :

- Scénario A :

$$\begin{aligned} & \left( \left( (0.7 \leq P(\text{faible}) \leq 1).X_{I,1} + (0 \leq P(\text{moyenne}) \leq 0.3).X_{I,2} \right) / 0.7.X^{M,1} \right. \\ & \quad \left. + \left( (0.4 \leq P(\text{faible}) \leq 1).X_{I,1} + (0 \leq P(\text{moyenne}) \leq 0.6).X_{I,2} \right) \right. \\ & \quad \left. / 0.3.X^{M,2} \right) \times (160/1.X^{J,1}) \end{aligned}$$

$$\begin{aligned} & \left( \left( (0.49 \leq P(\text{faible} \times 160) \leq 0.7).X_{I,1}.X^{M,1}.X^{J,1} \right. \right. \\ & \quad \left. + (0 \leq P(\text{moyenne} \times 160) \leq 0.21).X_{I,2}.X^{M,1}.X^{J,1} \right) \\ & \quad \left. + \left( (0.12 \leq P(\text{faible} \times 160) \leq 0.3).X_{I,1}.X^{M,2}.X^{J,1} \right. \right. \\ & \quad \left. \left. + (0 \leq P(\text{moyenne} \times 160) \leq 0.18).X_{I,2}.X^{M,2}.X^{J,1} \right) \right) \end{aligned}$$

- Scénario B :

$$\begin{aligned} & \left( (0.4 \leq \text{élevée} \leq 1).X_{K,1} + 0 \leq \text{très\_élevée} \leq 0.6).X_{K,2} \right) / 0.7.X^{M,1} \\ & \quad + \left( (0.2 \leq \text{élevée} \leq 1).X_{K,1} + 0 \leq \text{très\_élevée} \leq 0.8).X_{K,2} \right) / 0.3.X^{M,2} \\ & \quad \times (140/1.X^{L,1}) \end{aligned}$$

De la même manière que précédemment, on obtiendra les résultats suivants pour les deux scénarios A et B :

$$107.36 \leq UE(\varphi_{DS}(\text{Scénario A})) \leq 238.4$$

$$61.88 \leq UE(\varphi_{DS}(\text{Scénario B})) \leq 283.64$$

De manière logique, on trouve les mêmes résultats que pour les crédibilités et les possibilités obtenues lors de l'interprétation Demspter-Shafer. En effet, d'après le théorème faisant le lien entre possibilités et masses focales, « les éléments focaux sont totalement ordonnés par inclusion si et seulement si Bel et Pl sont respectivement une mesure de nécessité et de possibilité » (voir l'état de l'art).

### 7.2.2 Interprétation 2 : normalisation à un

Il s'agit d'une transformation de mesures de probabilités en mesures de possibilités. Une telle transformation s'effectue en normalisant chaque événement par Pmax. Après normalisation à un, on obtient les distributions de possibilité suivantes :

- Scénario A :

$$\left( (faible/1.X_{I,1} + moyenne/0.3.X_{I,2})/1.X_{M,1} + (faible/1.X_{I,1} + moyenne/0.6.X_{I,2})/\frac{3}{7}.X_{M,2} \right) \times (160/1.X^{J,1})$$

- Scénario B :

$$\left( (élevée/1.X_{K,1} + très_élevée/0.6.X_{K,2})/1.X_{M,1} + (élevée/1.X_{K,1} + très_élevée/0.8.X_{K,2})/\frac{3}{7}.X_{M,2} \right) \times (140/1.X^{L,1})$$

Après le développement des expressions ci-dessus, on obtient les formules développées suivantes :

- Scénario A :

$$\begin{aligned} & (\{faible\} \times 160)/1.X_{I,1}.X_{M,1}.X^{J,1} + (\{moyenne\} \times 160)/0.3.X_{I,2}.X_{M,1}.X^{J,1} \\ & + (\{faible\} \times 160)/\frac{3}{7}.X_{I,1}.X_{M,2}.X^{J,1} \\ & + (\{moyenne\} \times 160)/\frac{3}{7}.X_{I,2}.X_{M,2}.X^{J,1} \end{aligned}$$

- Scénario B :

$$\begin{aligned} & (\{élevée\} \times 140)/1.X_{K,1}.X_{M,1}.X^{L,1} + (\{très_élevée\} \times 140)/0.6.X_{K,2}.X_{M,1}.X^{L,1} \\ & + (\{élevée\} \times 140)/\frac{3}{7}.X_{K,1}.X_{M,2}.X^{L,1} \\ & + (\{très_élevée\} \times 140)/\frac{3}{7}.X_{K,2}.X_{M,2}.X^{L,1} \end{aligned}$$

On obtient alors les valeurs d'utilités possibles suivantes :



- Scénario A :  $(\{faible\} \times 160)/1 + (\{moyenne\} \times 160)/\frac{3}{7}$
- Scénario B :  $(\{élevée\} \times 140)/1 + (\{très\_élevée\} \times 140)/0.6$

On obtient alors une distribution de possibilité des utilités possibles :

- Scénario A :  $176/1 + 160/\frac{3}{7}$
- Scénario B :  $182/1 + 154/0.6$

Ensuite on pourra faire un minimax, un maximax, calculer une espérance avec une MDS, la nécessité, la crédibilité, etc.

Par exemple, on aura :

$$\varphi_{DS}(\text{Scénario A}) = \{176\}/\frac{4}{7} + \{176,160\}/\frac{3}{7}$$

$$betP(176) = \frac{4}{7} + \frac{3}{14} = \frac{11}{14} \approx 0.79$$

$$betP(160) = \frac{3}{14} \approx 0.21$$

$$UE(\varphi_{DS}(\text{Scénario A})) = \frac{11}{14} * 176 + \frac{3}{14} * 160 \approx 172.6$$

De la même manière :

$$\varphi_{DS}(\text{Scénario B}) = \{182\}/0.4 + \{182,154\}/0.6$$

$$betP(182) = 0.4 + 0.3 = 0.7$$

$$betP(154) = 0.3$$

$$UE(\varphi_{DS}(\text{Scénario B})) = 0.7 * 182 + 0.3 * 154 = 173.6$$

On pourra également évaluer les probabilités haute et basse de chaque scénario :

$$176 * \frac{4}{7} + 160 * 0 \leq UE(\varphi_{DS}(\text{Scénario A})) \leq 176 * 1 + 160 * \frac{3}{7}$$

$$100,57 \leq UE(\varphi_{DS}(\text{Scénario A})) \leq 244,57$$

$$182 * 0.4 + 154 * 0 \leq UE(\varphi_{DS}(\text{Scénario B})) \leq 182 * 1 + 154 * 0.6$$

$$72,8 \leq UE(\varphi_{DS}(\text{Scénario B})) \leq 274,4$$

Le tableau suivant synthétise les différents résultats obtenus en fonction des interprétations réalisées pour chaque scénario :

Tableau 30 : synthèse des différents résultats obtenus en fonction des interprétations réalisées pour chaque scénario

Interprétation	MDS (Mesure de Dempster-Shafer)				Normalisation à 1			
Indicateur \ scénario	Distribution	Utilité espérée	Utilité inférieure	Utilité supérieure	Distribution	Utilité espérée	Utilité inférieure	Utilité supérieure
Scénario A	$(\{faible\} \times 160) / 0.61 + (\{faible, moyenne\} \times 160) / 0.39$	172.88	107.36	238.4	$176 / 1 + 160 / \frac{3}{7}$	172.6	100,57	244.57
Scénario B	$(\{élevée\} \times 140) / 0.34 + (\{élevée, très\_élevée\} \times 140) / 0.66$	172.76	61.88	283.64	$182 / 1 + 154 / 0.6$	173.6	72.8	274.4

Le tableau suivant synthétise les décisions calculées en fonction des critères et des interprétations :

Tableau 31 : synthèse des décisions calculées en fonction des critères et des interprétations

Critère \ Décision	Décision pour l'interprétation MDS	Décision pour l'interprétation normalisation à 1
Utilité espérée	A	B
Utilité inférieure	A	A
Utilité supérieure	B	B

Dans l'étude de cas que nous venons de présenter, la normalisation à 1 donne les résultats les plus précis. Entre le scénario A et le scénario B, la différence principale réside dans le plus grand équilibre entre les deux éventualités « élevée » et « très\_élevée », ce qui fait que l'éventail des espérances possibles est plus large. Ainsi, si les utilités des deux scénarios sont équivalentes du point de vue de l'espérance, notre approche permet d'identifier et de calculer l'éventail des utilités possibles des deux scénarios. Le choix « prudent » conduira à choisir le scénario A tandis que le choix « risqué » conduira à choisir le scénario B. Ces résultats sont cohérents avec l'approche de prise de décision dans le risque, à savoir un minimax de 160 (scénario A) et un maximin de 182 (scénario B).

### 7.3 Conclusion : deux études de cas pour l'aide à la décision dans des exploitations agricoles mettant en œuvre les concepts précédemment définis

Nous venons de présenter deux études de cas :

- La première présente l'intérêt de distinguer probabilités et possibilités dans certains cas. Une telle approche distingue les probabilités devenant de plus en plus « certaines » au fur et à mesure que les informations s'accumulent et des possibilités qui restent inchangées.
- La seconde propose de traiter le cas de distributions de probabilité et possibilités imbriquées avec des interprétations particulières transformant les deux types de distributions dans un formalisme commun.

Dans le prochain chapitre, nous proposerons une architecture distribuée pour la validation, la mise en œuvre et l'opérationnalisation d'applications basées sur les concepts précédemment définis.

## 8 Chapitre 8 : opérationnalisation : un modèle architectural pour la mise en œuvre d'applications dédiées aux systèmes socio-environnementaux

Dans ce chapitre, nous présentons un modèle architectural pour la validation, la mise en œuvre et l'opérationnalisation d'applications basées sur les concepts précédemment définis. Nous proposons une architecture orientée service qui supporte des propriétés de découplage, de réutilisation, de composabilité, de modularité et d'encapsulation. Une telle architecture permet à l'entreprise d'exposer des services à ses clients et partenaires, ou d'autres parties de l'organisation, en vue de leur collaboration.

Un exemple d'application que nous allons présenter est une plateforme orientée services ayant pour sources de données des capteurs situés dans l'exploitation agricole et des bases de données internes ou externes à l'exploitation. Ces données disponibles potentiellement mises à jour en continu alimentent la décision et le conseil à tout instant, le calcul étant fait « au mieux » sur les données disponibles en tenant compte de leurs imperfections.

### 8.1 Introduction : architectures orientées services et services Web

Une Architecture Orientée Service<sup>43</sup> est une architecture distribuée mettant en œuvre des composants logiciels appelés services. Ces derniers sont reliés entre eux par des flux de données et sont réutilisables. Une telle architecture est caractérisée par un couplage faible avec les autres applications par l'utilisation d'une couche d'interface interopérable, en général des Services Web<sup>44</sup>, ainsi que par l'utilisation d'un format d'échange favorisant

---

<sup>43</sup> Dans la suite de ce document, nous désignerons les Architectures Orientées Service par leur acronyme anglais, SOA (pour « Service Oriented Architecture »)

<sup>44</sup> Dans la suite de ce document, nous désignerons les Services Web par leur acronyme anglais, WS (pour « Web Service »). Les services Web sont des entités informatiques autonomes qui permettent la communication et l'échange de données entre des applications et

l'interopérabilité tel que le langage XML<sup>45</sup> [W3C, 2008] ou le langage JSON<sup>46</sup> [ECMA, 2013]. L'utilisation de ces standards permet notamment d'écrire des WS dans plusieurs langages et de les utiliser sur des systèmes d'exploitation différents.

Nous proposons une application dédiée à la modélisation de systèmes socio-environnementaux. Celle-ci est basée sur le style d'architecture répartis appelé REST<sup>47</sup>. Nous présentons ces principes ainsi que les avantages qui nous ont conduit à choisir ce style d'architecture ci-dessous [Fielding, 2000].

- Identification des ressources : les éléments et états gérés dans l'application sont considérés comme des ressources abstraites identifiées grâce à un système universel d'identification : les URI<sup>48</sup>. Leur représentation est accessible sous un format léger lisible par n'importe quelle application (XML ou JSON).
- Interface uniforme : navigation et manipulation au sein des ressources hypermedia à travers des représentations prédéfinies (schémas JSON ou XML) et reposant sur un

---

des systèmes hétérogènes, dans des environnements distribués et mises à disposition sur un réseau.

<sup>45</sup> Acronyme de eXtended Markup Language. XML est un métalangage à base de balises non prédéfinies conçu pour structurer et décrire des données. Il s'agit d'un standard du W3C. L'articulation des WS autour d'un tel langage leur procure l'avantage d'être non propriétaires et multi-plate-formes.

<sup>46</sup> Acronyme de JavaScript Object Notation. JSON est un format textuel léger d'échange de données complètement indépendant de tout langage. Il est basé sur un sous-ensemble du langage de programmation JavaScript. Il s'agit d'un standard de l'ECMA. Ce format est plus léger que XML et très répandu notamment présent dans les BDD No-SQL orientées document telles que MongoDB.

<sup>47</sup> REST est l'acronyme de Representational State Transfer. REST est un style d'architecture client-serveur pour systèmes hypermedia basé sur un certain nombre de principes qui décrivent la manière dont les informations sont définies et circulent sur le réseau.

<sup>48</sup> URI est l'acronyme de Uniform Resource Identifier. Une URI est une chaîne de caractères identifiant une ressource sur un réseau (par exemple une ressource Web) physique ou abstraite.

nombre limité d'opérations (GET, POST, PUT et DELETE). La nature hypermédia de l'application permet d'accéder aux états suivants de l'application par inspection de la ressource courante.

- Interactions sans état (« stateless ») : l'état de l'application est géré côté client et non pas côté serveur. Chaque requête du client vers le serveur contient toutes les informations nécessaires pour que cette demande soit comprise, et elle ne peut tirer profit d'aucun contexte stocké sur le serveur. Ce principe a pour conséquences une plus grande indépendance entre le client et le serveur, la non-nécessité d'une connexion permanente entre ces derniers (i.e. pas de session client à maintenir sur un serveur particulier), et enfin la possibilité de répartir des requêtes sur plusieurs serveurs, d'où une plus grande tolérance aux pannes.

Aux principes liés à l'identification, à l'interface et aux interactions des ressources s'ajoutent les principes architecturaux suivants :

- Cache : l'ajout de contraintes de cache offre la possibilité d'éliminer certaines interactions répétitives. Les messages étant auto-descriptifs, les informations nécessaires sur la fraîcheur, la péremption de la ressource sont contenues dans le message lui-même (amélioration des performances notamment lors d'une brusque montée en charge du serveur).
- Système en couches : des intermédiaires tels que des serveurs mandataires, des passerelles et des pare-feux peuvent être introduits à divers endroits dans la communication sans pour autant changer les interfaces entre les composants, les messages étant auto-descriptifs.
- Code à la demande (facultatif) : les clients peuvent exécuter des scripts obtenus depuis le serveur, ce qui simplifie l'implémentation des clients en réduisant le nombre de fonctionnalités qu'ils doivent mettre en œuvre par défaut.

## 8.2 Processus de conception d'une application innovante dédiée aux systèmes socio-environnementaux

Nous proposons une application basée sur une architecture orientée service permettant à l'utilisateur (par exemple un spécialiste du domaine agricole) de concevoir ses propres applications agricoles innovantes à base du méta-modèle systémique proposé dans nos

contributions et de WS. Le processus de conception de telles applications comprend des étapes suivantes :

**(1) Information** : des WS implémentant des opérations de type CRUD<sup>49</sup>, pour l'interrogation et le recensement de données provenant de multiples sources.

**(2) Modélisations structurelle et fonctionnelle** : l'expert du domaine formalise son propre modèle du domaine<sup>50</sup> (modélisation des blocks, des classes, de l'imperfection) à partir de modèles génériques du domaine qui sont accessibles sur le Web via des WS. Ensuite, il modélise les processus qui régissent l'ensemble des éléments modélisés via des modèles fonctionnels qui coordonnent les éléments précédemment définis grâce aux flux d'éléments qui circulent entre les blocks.

**(3) Opérationnalisation** du modèle fonctionnel précédemment établi, qui s'appuie sur :

- Le modèle fonctionnel précédemment défini,
- Le modèle des équations qui décrit les processus régissant le comportement des blocks (relations entre les entrées et les sorties, paramètres),
- Le chargement des données et le mapping entre le modèle métier précédemment établi et les données extraites de différentes sources de données dynamiques ou persistantes (e.g. SQL, XML, JSON) via les services de fourniture de données et paramétrage de leurs imperfections.

**(4) Simulation** du modèle via un moteur d'orchestration des services afin de simuler les chaînes fonctionnelles de calcul prenant en compte l'imperfection des données.

---

<sup>49</sup> CRUD est l'acronyme de « create, read (ou retrieve), update, delete. CRUD désigne les quatre opérations de base pour la persistance des données (bases de données).

<sup>50</sup> On appelle modèle du domaine un modèle quelconque déjà existant disponible par exemple sur le Web. Le modèle structurel est quant à lui le modèle créé et utilisé par l'expert du domaine dans l'application à partir de modèles du domaine. Dans un contexte objet, nous qualifierions ces notions de « classes métier ».

## 8.3 Une architecture orientée service dédiée aux systèmes socio-environnementaux

L'architecture que nous proposons se compose d'une interface utilisateur et de services externes.

### 8.3.1 Services externes

Les éléments applicatifs suivants sont des services externes. Ils sont répartis en deux catégories :

Catégorie « domaine » :

- Services de fourniture du modèle de l'imperfection : fourniture de différents types d'imperfections basées sur le type d'imperfection et le degré de variation associé. Les paramètres fournis au modèle du domaine sont des paramètres par défaut.
- Services de fourniture de modèles du domaine contenant les éléments de sens ainsi que leurs relations [Dantan et al., 2012 (1)] [Dantan et al., 2014 (1)], interfaces permettant de récupérer des données relatives à des modèles du domaine disponibles sur Internet, e.g. AGROVOC [FAO, 2015].
- Services de fourniture des modèles structurel et fonctionnel et opérationnel réalisés par l'utilisateur et stockés dans un répertoire de modèles.

Catégorie « données » :

- Services de fourniture des données à partir des descriptions de sources et d'objets [Dantan et al., 2014 (1)] fournies par les modèles structurel et opérationnel réalisés par l'utilisateur (opérations de type CRUD).
- Services utilitaires génériques : fourniture d'opérations élémentaires implémentées par les équations : opérations algébriques, relations mathématiques entre variables, paramètres et résultats de modèles analytiques, etc.

### 8.3.2 Interface utilisateur

L'interface utilisateur permet à l'expert du domaine de modéliser le système socio-environnemental voulu, de séquencer les différentes étapes du modèle via les modèles fonctionnel et opérationnel du profil SysML, et de générer le code (classes,



accesseurs/mutateurs génériques définis automatiquement, instances, imperfections) qui exécute ces séquences dans un environnement d'exécution. Ce module coordonne l'état des éléments modélisés à tout instant. Il est implémenté avec le langage orienté objet Python pour l'instanciation dynamique des modèles abstraits et des modèles concrets implémentant les blocs (modèle structurel) et les flux (modèle fonctionnel) qui composent le type de système socio-environnemental modélisé (par exemple une exploitation agricole) via les services de fourniture de modèles du domaine, de l'imperfection, des données et les services utilitaires génériques.

Remarque : l'infrastructure proposée pourra être intégrée sur une plate-forme orientée cloud computing. En effet, les connections à un Cloud sont en général réalisées via des WS.

Un aspect important de la modélisation de système est l'échange de données entre outils. Un format standard d'échange de données XMI (XML Metadata Interchange) a été développé<sup>51</sup>. Cependant, les implémentations XMI de SysML variant selon les éditeurs, nous avons décidé d'élaborer notre propre schéma XML du profil de métamodèle SysML. Ainsi, les modèles SysML stockés dans un entrepôt de données (repository) peuvent tout de même être importés/exportés par d'autres applications. Une interface réalisant le mapping entre les sémantiques de schémas sera éventuellement à développer dans le futur. Nous avons choisi le format de données semi-structurées JSON pour modéliser les données qui circulent sur le réseau.

### 8.3.3 Illustration de l'application proposée

Le schéma suivant illustre l'application basée sur l'architecture orientée service que nous proposons :

---

<sup>51</sup> Disponible à la page suivante : « Documents Associated with Systems Modeling Language (SysML), Version 1.4 » : <http://www.omg.org/spec/SysML/1.4/>

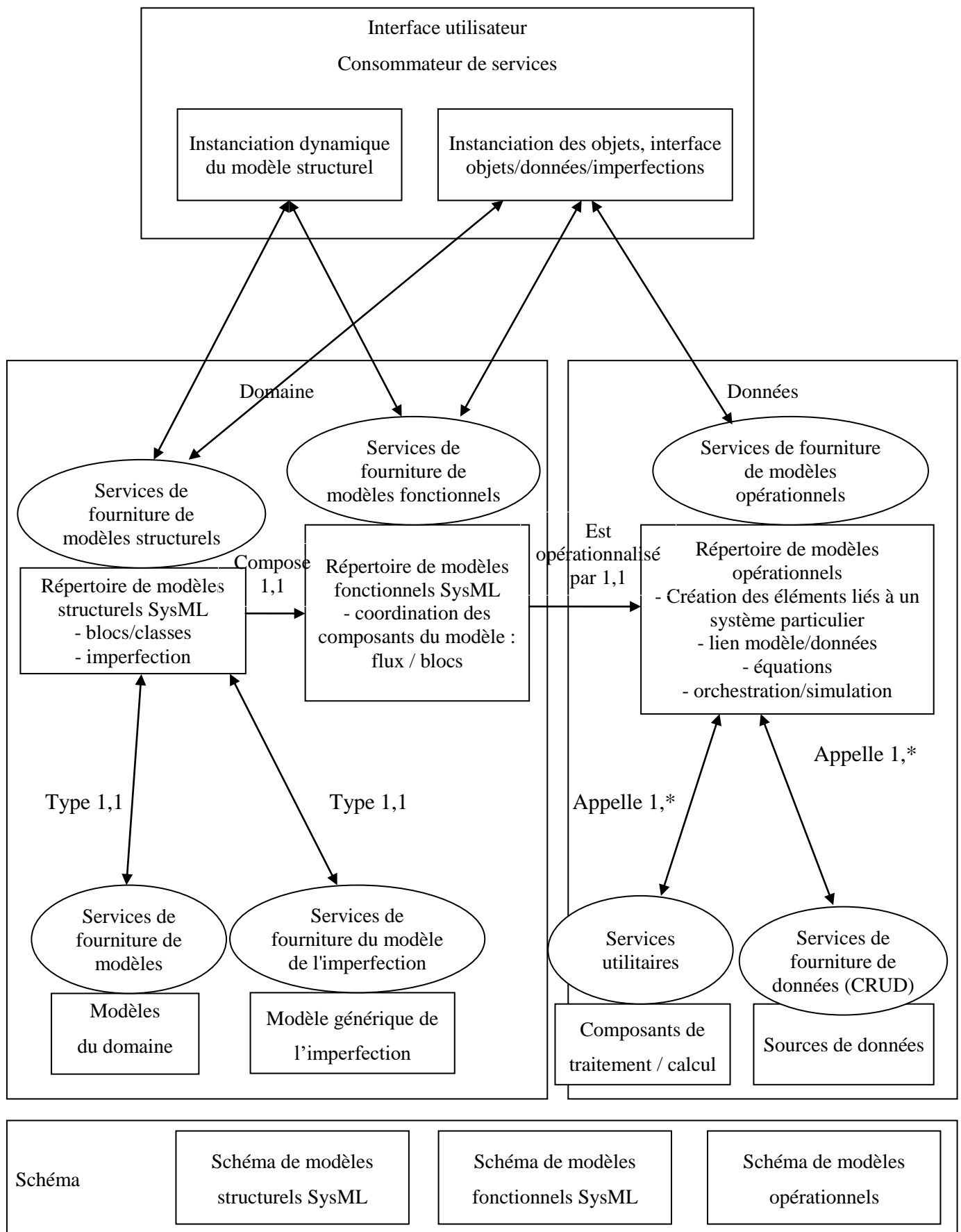


Figure 23 : architecture de l'application proposée

### 8.3.4 Index des services exposés

L'application SOA que nous proposons s'appuie sur le style d'architecture REST. Nous avons identifié les différentes ressources (URI) qui seront exposées en tant que services, les chemins d'accès associés et les possibilités des quatre opérations de base disponibles.

- La méthode GET est disponible sur l'ensemble des ressources.
- Les méta-modèles structurel et fonctionnel ne peuvent pas être modifiés (sauf les flux dont on peut préciser les blocks qui y sont liés).
- La méthode POST permet d'ajouter des ressources filles (classes) à un conteneur. Les méthodes PUT et DELETE permettent de mettre à jour / supprimer des ressources filles (instances).

Les tableaux suivants illustrent ces associations pour chacun des modèles exposés :

Tableau 32 : ressources du modèle structurel

<b>Operations Resources</b>	<b>GET (retrieve)</b>	<b>PUT (create, update)</b>	<b>POST (append)</b>	<b>DELETE (delete)</b>
Structural model / blocks	O	N	N	N
Structural model / blocks / {name}	O	O	N	O
Structural model / blocks / {name} / property	O	N	N	N
Structural model / blocks / {name} / properties / {name}	O	O	N	O
Structural model / imperfection /	O	N	N	N
Structural model / imperfection / {type}	O	N	N	N
Structural model / imperfection / {type} / variationlist	O	N	N	N
Structural model / imperfection / {type} / variationlist / {nameslist}	O	N	N	N

Tableau 33 : ressources du modèle fonctionnel

<b>Operations Resources</b>	<b>GET (retrieve)</b>	<b>PUT (create, update)</b>	<b>POST (append)</b>	<b>DELETE (delete)</b>
Functional model / flows	O	N	N	N
Functional model / flows / {name}	O	O	N	O

<b>Operations Resources</b>	<b>GET (retrieve)</b>	<b>PUT (create, update)</b>	<b>POST (append)</b>	<b>DELETE (delete)</b>
Functional model / flow / fromblock   toblock   block /	O	N	O	N
Functional model / flow / fromblock   toblock   block / {blockname}	O	O	N	O
Functional model / flow / block / {blockname} / property	O	N	O	N
Functional model / flow / block / {blockname} / property / {propertyname}	O	O	N	O

Tableau 34 : ressources du modèle opérationnel

<b>Operations Resources</b>	<b>GET (retrieve)</b>	<b>PUT (create, update)</b>	<b>POST (append)</b>	<b>DELETE (delete)</b>
Operational model / block	O	N	O	N
Operational model / block / {id}	O	O	N	O
Operational model / block / {id} / property / {id}	O	N	O	N
Operational model / block / {id} / property / {id}	O	O	N	O
Operational model / block / {id} / property / {id} / imperfection	O	N	O	N
Operational model / block / {id} / property / {id} / imperfection / {type}	O	O	N	O
Operational model / block / {id} / property / {id} / imperfection / {type} / variationlist	O	N	O	N
Operational model / block / {id} / property / {id} / imperfection / {type} / variationlist / {values}	O	O	N	O
Operational model / block / {id} / equation	O	N	O	N
Operational model / block / {id} / equation / {id} /	O	O	N	O
Operational model / block / {id} / equation / {id} / operand1   operand2	O	N	O	N
Operational model / block / {id} / equation / {id} / operand1   operand2 / {fromblockid}	O	O	N	O

<b>Operations Resources</b>	<b>GET (retrieve)</b>	<b>PUT (create, update)</b>	<b>POST (append)</b>	<b>DELETE (delete)</b>
Operational model / block / {id} / equation / {id} / result	O	N	O	N
Operational model / block / {id} / equation / {id} / result / { toblockid}	O	O	N	O

## 8.4 Fonctions proposées

### 8.4.1 Une interface pour la gestion de l'imperfection des données

Nous avons conçu et développé une API<sup>52</sup> de gestion de l'imperfection des données. Cette API met à disposition les formalismes de représentation et de calculs sur des quantités incertaines que nous avons proposées dans nos contributions, à savoir :

- Trapezoid : sous-ensemble flou de valeurs possibles, représenté par un « trapèze ».
- Normal : VA de loi normale.
- FuzzyDiscrete : distribution de possibilité discrète.
- ProbaDiscrete : distribution de probabilité discrète.
- ProbaNormalCollection : collection possibiliste discrète de sous-ensembles flous de valeurs possibles, représentés par des « trapèzes ».
- FuzzyTrapezoidCollection : collection probabiliste discrète de VA de loi normale.

Dans l'approche que nous proposons, les éléments d'imperfection sont indépendants (orthogonaux) de l'intention, i.e. ils peuvent être associés a posteriori aux éléments du modèle du domaine (classes métiers et/ou attributs). Concrètement, une fois le schéma de données défini, nous superposons à celui-ci les éléments d'imperfection (exemples : valeurs d'attributs ou existence de classes plus ou moins possibles ou probables) à l'aide d'annotations dont les paramètres spécifient les méta-informations permettant d'interpréter une base classique nette comme une base floue / probabiliste. Ainsi la gestion de l'imperfection est indépendante du

---

<sup>52</sup> API est l'acronyme d'Application Programming Interface (interface de programmation). Il s'agit d'un ensemble de fonctions qui permet à un logiciel (consommateur) d'accéder à des services d'autres logiciels (fournisseur, par exemple via une bibliothèque logicielle ou un service web)

schéma de données. L'intérêt de cette indépendance est que cela offre la possibilité aux experts du domaine qui réalisent les modèles fonctionnels du système de travailler indépendamment des experts du domaine qui décrivent les sources. De plus, une telle approche permet la réutilisation des codes existants.

#### 8.4.2 Une interface pour l'introspection des fonctions et classes

Nous proposons également une API de découverte dynamique des classes et de leurs propriétés lors de leur création. Les paramètres et le nombre d'appels des fonctions sont également affichés. Ces fonctions ont été implémentées via des annotations Python (« `traceClass` » et « `traceFunction` »).

Le but de notre démonstrateur n'est pas de réaliser un nouvel éditeur SysML mais de réaliser des interfaces (via des services) suffisamment génériques pour étendre les outils existants avec le profil SysML proposé, moyennant bien sûr l'existence des interfaces de communication nécessaires dans l'outil cible.

### 8.5 Conclusion : une nouvelle approche de conception d'applications innovantes dédiées aux systèmes socio-environnementaux

Nous avons proposé une application basée sur une architecture orientée service REST qui implémente : information, modélisation, opérationnalisation et simulation. Les services dédiés à ces processus sont exposés à une organisation, en vue de la collaboration des acteurs qui la composent. Le code source de cette application est annexé à notre thèse.

## Cinquième partie : conclusion

# Conclusion générale et perspectives

## Synthèse des travaux réalisés

Les résultats de notre thèse s'articulent autour des quatre contributions majeures suivantes :

- **Un méta-modèle systémique pour la modélisation, la simulation et l'« optimisation durable » des systèmes socio-environnementaux.** Ce méta-modèle est constitué de quatre modèles : le modèle du domaine représenté par un ensemble de concepts pour la modélisation du domaine de l'agriculture, le modèle structurel qui définit des entités ainsi que des types d'attributs génériques exprimant différents types d'imperfection, le modèle fonctionnel qui définit des systèmes socio-environnementaux génériques via des flux et des blocs fonctionnels et enfin le modèle des équations qui opérationnalise les modèles précédents. Nous avons par ailleurs proposé une correspondance naturelle entre les concepts proposés et les notations classiques du langage SysML (profil), tout en apportant la rigueur sémantique indispensable.
- **Une approche formelle pour la représentation et la combinaison de données imparfaites.** Nous avons défini un formalisme unifié pour la représentation et la manipulation de quantités imparfaites, qui peuvent prendre un nombre fini ou infini de valeurs possibles et probables, dont nous avons défini la nécessité dans la contribution précédente. Nous avons obtenu une véritable structure algébrique qui exploite l'enchaînement des calculs sur de telles quantités avec des propriétés similaires à  $\mathbb{R}$ . Nous avons enfin présenté une structure de données orientée objet adaptée au stockage des quantités « nettes » et des quantités imparfaites, ainsi que des opérateurs génériques d'agrégation.
- **Un modèle de décision en contexte incertain.** Nous avons défini un certain nombre de transformations particulières que nécessitent les opérations algébriques sur des quantités imparfaites pour ensuite les agréger en un type unique de distribution, en vue de prendre des décisions.
- **Une opérationnalisation des concepts précédemment définis.** Dans un premier temps, nous avons présenté deux études de cas qui traitent différents problèmes d'aide



à la décision sur la période de semis dans une exploitation agricole. Ensuite, nous avons proposé une architecture orientée service basée sur les concepts précédemment définis, et plus particulièrement en rapport avec les différents types d'imperfections qui peuvent entacher des données.

## Apports au domaine de l'ingénierie des systèmes socio-environnementaux

Les recherches que nous avons menées dans cette thèse contribuent plus particulièrement au domaine de l'ingénierie des systèmes de décision en univers incertain, dans un contexte de grandes masses de données traitées dynamiquement par des modèles complexes résultant de l'assemblage de modèles élémentaires.

Nous proposons donc une approche unifiant des apports des différents domaines concernés, tirant parti des avantages de chacun d'entre eux et les faisant opérer en synergie : la modélisation des systèmes via un méta-modèle pour les systèmes socio-environnementaux, la représentation générique de l'imperfection, la prise de décision dans ce contexte et enfin la mise en œuvre de ces concepts via une architecture orientée service pour la modélisation coopérative et la simulation en environnement dynamique.

Nos contributions ont apporté un gain notable au vu des limites des approches existantes que nous avons identifiées dans l'état de l'art :

- L'intégration des notions d'incertitude sur les données, plus particulièrement floues et probabilistes à SysML, langage standard de modélisation des systèmes.
- La clarification de la syntaxe graphique dans SysML, notamment en ce qui concerne les liens entre les diagrammes structurel et fonctionnel, le diagramme paramétrique et la nature des ports et des valeurs. Le pré-paramétrage des relations entre blocs des diagrammes avec des notions d'agréats constitue un moyen intéressant d'automatiser les calculs après modélisation.
- La modélisation unifiée des différentes imperfections entachant les informations, sachant que, en raison du volume, de l'hétérogénéité et de la vitesse/fréquence de traitement des données (« big data ») dans le contexte des systèmes socio-environnementaux, les approches probabilistes et la méthode de Monte-Carlo ne suffisent pas toujours. Notre approche consiste à envisager des scénarios croisés selon

différentes hypothèses incertaines, remplaçant  $m$  calculs numériques sur  $n$  hypothèses par un seul calcul formel, tout en s'intégrant au méta-modèle systémique que nous avons proposé.

- Une démarche très simple de traitement de données imparfaites. A l'inverse d'autres approches qui unifient différents types d'imperfections, telles que la théorie de Dempster-Shafer, nous proposons des combinaisons rigoureuses et opérationnelles d'informations qui se résument à une démarche très simple de traitement des données : calculs rigoureux, choix d'une interprétation particulière par le scientifique et enfin agrégation en vue de la prise de décisions. Notre approche ne se base donc pas sur une unique interprétation de la fusion de données dont les imperfections sont hétérogènes, issues de différentes théories de l'incertain, mais nous fournissons un cadre formel permettant de propager de manière rigoureuse les différents types d'incertitudes qui pourraient être interdépendants, sur lesquels une interprétation particulière de fusion de données pourra être appliquée.
- Une structure algébrique qui exploite l'enchaînement des calculs sur de telles quantités avec des propriétés similaires à  $\mathbb{R}$ , ce que ne permettent pas les approches classiques basées sur les quantités floues proposées dans la littérature. Les avantages pratiques permettent de construire de manière rigoureuse une situation avec des dépendances entre imprécisions avec des disjonctions d'expressions. De plus, les problèmes de fausses extensions de l'univers des possibles sont évités, de même que les termes croisés, lors de calculs sur des quantités mixtes, sont évalués à la fin des calculs, de telle sorte que les données sont agrégées indépendamment de la méthode spécifique de combinaison possibilités/probabilités choisie et que les calculs demeurent rigoureux.
- Un ensemble d'opérateurs génériques d'agrégation pour évaluer et synthétiser les incertitudes et les différents scénarios possibles pour la prise de décision. De nombreuses interprétations particulières et opérateurs sont disponibles dans la littérature. L'avantage de notre approche est que le formalisme unifié proposé est compatible avec ces autres interprétations qu'il convient d'implémenter en fonction des besoins.
- Une architecture orientée service réalisée sur la base des concepts précédemment définis. Dans l'application proposée, la gestion de l'imperfection est indépendante du

schéma de données i.e. ils peuvent être associés a posteriori aux éléments du modèle du domaine (classes métiers et/ou attributs). L'intérêt de cette indépendance est d'offrir la possibilité aux experts du domaine qui réalisent les modèles fonctionnels du système de travailler indépendamment des experts du domaine qui décrivent les sources. Enfin, une telle approche permet la réutilisation des codes existants.

## Perspectives

La suite de nos travaux, de par la grande extensibilité de notre approche, comporte de nombreuses perspectives scientifiques, techniques et applicatives :

- De nouveaux modèles de décision intégrant l'imperfection. D'une part, une piste intéressante est l'implémentation d'autres lois de probabilités et d'autres interprétations particulières utilisées dans d'autres domaines experts que celui étudié. Nous avons également observé la non-stabilité de la plupart des lois de probabilité continues classiques. A notre avis, il est pertinent de poursuivre cette étude sur la stabilité des lois par opérations algébriques classiques, notamment en approximant certains produits de lois continues par des lois discrètes.
- Une plateforme pour la décision en contexte incertain destinée aux experts du domaine : l'intégration de l'extension de SysML à des éditeurs du marché, le maintien de la conformité de notre extension aux évolutions à venir du langage SysML, concernant notamment les progrès en terme de standardisation et enfin l'intégration d'autres modèles agronomiques existants à la plateforme.
- Un support au domaine de l'agriculture et d'une manière plus générale à l'optimisation durable de la planète dans un contexte de dérèglement climatique et dans le cadre de nouvelles méthodes d'optimisation de gestion des ressources.

# Publications et conférences

## Publication réalisée dans le cadre de la thèse

- Dantan, J., Pollet, Y., Taïbi, S. 2015. Combination of imperfect data in fuzzy and probabilistic extension classes. *Journal of Environmental Accounting and Management*. 3 (2), 123-150. DOI: 10.5890/JEAM.2015.06.004.

## Communications avec actes dans des conférences d'audience internationale réalisées dans le cadre de la thèse

- Dantan, J., Pollet, Y., Taïbi, S. 2016. A formal approach for the representation and the combination of imperfect data. *Agrostat 2016 congress*, March 21-24 2016, Lausanne, Switzerland.
- Dantan, J., Pollet, Y., Taïbi, S., 2015. A systemic meta-model for socio-environmental systems. In *proceedings of the Sixth International Conference on Complex Systems Design & Management, CSD&M 2015*. Editors: Auvray, G., Bocquet, J.-C., Bonjour, E., Krob, D. (Eds.). November 23-25, 2015. P. 307. Paris, France.
- Dantan, J., Pollet, Y., Taïbi, S. 2015. A formal model to compute uncertain continuous data. In *proceedings of CS-DC'15 World e-conference (Complex Systems Digital Campus) UNITWIN/UNESCO*. September 28 - October 2, 2015.
- Dantan, J., Pollet, Y., Taïbi, S. 2014. Taking account of uncertain, imprecise and incomplete data in sustainability assessments in agriculture. In *proceedings of Computational Science and Its Applications - ICCSA 2014 - 14th International Conference, Part III, Lecture Notes in Computer Science LNCS 8581, ISBN 978-3-319-09149-5*, pp. 625–639. Springer International Publishing Switzerland. Guimarães, Portugal, June 30 - July 3, 2014.
- Dantan, J., Pollet, Y., Taïbi, S. 2014. A goal-oriented meta-model for scientific research. In *proceedings of Computational Science and Its Applications - ICCSA 2014 - 14th International Conference, Part V, Lecture Notes in Computer Science LNCS*

8583, ISBN 978-3-319-09155-6, pp. 762–774. Springer International Publishing Switzerland. Guimarães, Portugal, June 30 - July 3, 2014.

- Dantan, J., Pollet, Y., Taïbi, S. 2013. The G.O.A.L. Approach. In proceedings of ENASE International Conference on Evaluation of Novel Approaches to Software Engineering. July 4-6, 2013, pp. 173-180. Angers, France.
- Dantan, J., Pollet, Y., Taïbi, S., 2012. Semantic indexation of Web services for collaborative expert activities. In proceedings of IADIS International Conference Information Systems. March 10-12, 2012, pp. 57-64, Berlin, Germany.
- Dantan, J., Pollet, Y., Taïbi, S., 2012. A KDD Process to retrieve and aggregate data from relational databases. In proceedings of IADIS International Conference Information Systems. March 10-12, 2012, pp. 443-445, Berlin, Germany.

## Autres publications et conférences

### Publication

- Taïbi-Hassani, S., Thoisy-Dur, J-C., Lepelletier, P., Bodin, J., Bennegadi-Laurent, N., Bessoule, J-J., Bispo, A., Bodilis, J., Chaussod, R., Cheviron, N., Cortet, J., Criquet, S., Dantan, J Dequiedt, , A., Faure, O., Gangneux, C., Harris-Hellal J., Hedde, M., Hitmi, A., Le Guedard, M., Legras, M., Pérès, G., Repinçay, Rougé, L., C., Ruiz, N., Trinsoutrot-Gattin, I., Villenave, C. 2013. Démarche statistique pour la sélection des indicateurs par Random Forests pour la surveillance de la qualité des sols. Etude et Gestion des Sols. 20 (2), 127-136.

### Communications

- Taïbi, S., Dantan, J., 2016. The advent of Big Data in Agriculture. Agrostat 2016 congress, March 21-24 2016, Lausanne, Switzerland.
- Taïbi, S., Lepelletier, P., Dantan J., Thoisy-Dur, J.-C., Bodin, J. 2014. A Statistical approach for soil monitoring, risk assessment and soil characterization, e-Kickoff ICCSA'14 , Complex Systems Digital Campus, UNITWIN-UNESCO June 23-26th 2014.
- Taïbi, S., Rougé, L., Thoisy-Dur, J.-C., Bodin, J., Lepelletier, P., Dantan, J., Pérès, G., Grand, C. , Bispo, A., 2012. Gestion et traitement des données du programme. Approche statistique de sélection d'Indicateurs et de biomarqueurs dans la surveillance

de la qualité des sols et l'évaluation des risques. Journées Techniques Nationales, Bioindicateurs pour la caractérisation des sols, ADEME, Paris, 10 pages, 16 octobre 2012.

- Taïbi, S., Thoisy-Dur, J.-C., Bodin, J., Rougé, L., Dantan, J., Lepelletier, P., Michaud, A., Houot, S., Pérès, G., Bispo, A. 2013. A statistical approach to assess soil biodiversity and biological activity responses to repeated organic amendment applications in cultivated soils - Relationships with soil functions. RAMIRAN, 15th international conference. 3-5 June, Versailles, France.
- Bodin, J., Dur, J.-C., Rougé, L., Dantan, J., Lepelletier, P., Grand, C., Pérès, G., Bispo, A. Taïbi, S., 2013. Soil bioindicators to assess soil biodiversity and activity responses to land-use practices. Final results of the research project Bioindicators. International Interdisciplinary Conference on Land Use and Water Quality: Reducing Effects of Agriculture. 10-13 June, The Hague, The Netherlands.
- Thoisy-Dur, J.-C., Lepelletier, P., Taïbi, S., Rougé, L., Dantan, J., Pérès, G., Grand, C. and Bispo, A. 2012. Statistical approach to select soil bioindicators for soil monitoring, risk assessment and soil characterization. Results from the French national Program Bioindicators. 6th SETAC World Congress/SETAC Europe 22nd Annual Meeting. Mai 20-24, Berlin, Allemagne.
- Taïbi, S., Dur, J.-C., Lepelletier, P., Rougé, L., Dantan, J., Bispo, A., Grand, C., Pérès, G. 2012. Approche statistique de sélection d'Indicateurs et de Biomarqueurs dans la surveillance de la qualité des sols et l'évaluation des risques. Résultats du programme national ADEME Bioindicateurs. Mars 2012, JES Versailles, France.
- Taïbi, S., Thoisy-Dur, J.-C., Lepelletier, P., Rougé, L., Dantan, J., Bispo, A., Grand, C., Pérès, G. 2012. Approche statistique de sélection d'indicateurs et de biomarqueurs dans la surveillance de la qualité des sols et l'évaluation des risques. Résultats du programme national ADEME Bioindicateurs. 11èmes Journées d'Etude des Sols. March, 19-23, Versailles.
- Thoisy-Dur, J.-C., Lepelletier, P., Taïbi, S., Rougé, L., Dantan, J., Pérès, G., Grand, C., Bispo, A. 2012. Statistical approach to select soil bioindicators for soil monitoring, risk assessment and soil characterization. Results from the French national Program Bioindicators SETAC-Europe, Berlin.

- Bodin J., Taïbi, S., Thoisy-Dur J.-C., Dantan J., Lepelletier, P., Rougé L. 2013. Bioindicateurs de la qualité des sols. Démarche d'analyse globale. Rapport d'activités. ADEME- Esitpa Fév. 2013.

#### Autres productions

- Bodin J., Taïbi, S., Thoisy-Dur J.-C., Dantan J., Lepelletier, P., Rougé L. 2013. Bioindicateurs de la qualité des sols. Démarche d'analyse globale. Rapport d'activités. ADEME- Esitpa Fév. 2013.





## 9 Bibliographie

[Abdellaoui et Gonzales, 2009] Abdellaoui M., Gonzales C., 2009. *Multiattribute Utility Theory*. Chapter in *Decision Making Process. Concepts and Methods*, D. Bouyssou, D. Dubois, M. Pirlot, H. Prade (Eds.), pp. 555-588, (ISTE -- Wiley), (ISBN: 978-1-84821-116-2) (2009)

[ACTA, ICTA, INRA , 2004] ACTA, ICTA, INRA, 2004. *Diaphyt : un outil de diagnostic sur les effets non intentionnels des produits phytopharmaceutiques à l'échelle de l'exploitation agricole*. [En ligne], consulté le 01/03/2016, [http://www.acta.asso.fr/formations/formations-ita/fo/formation/detail/diaphyt\\_initiation\\_a\\_lutilisation\\_dun\\_logiciel\\_sur\\_les\\_bonnes\\_pratiques\\_dutilisation\\_des\\_pro-2.html](http://www.acta.asso.fr/formations/formations-ita/fo/formation/detail/diaphyt_initiation_a_lutilisation_dun_logiciel_sur_les_bonnes_pratiques_dutilisation_des_pro-2.html)

[Araujo et al., 1999] Araujo C., Araujo Bonjean C., Arcand J-L, 1999. *Capital humain, productivité agricole, et travail féminin : variables latentes et séparabilité dans les modèles de ménage*. [En ligne], No 199912, Working Papers, CERDI, consulté le 01/03/2016, <http://cerdi.org/uploads/ed/1999/1999.12.pdf>

[Attonaty, 2010] Attonaty, J. M., 2010. *Olympe : simulateur technico-économique des exploitations agricoles*. [En ligne], INRA, consulté le 01/03/2016, <http://www.olympe-project.net/>

[Baize, 1995] Baize D., 1995. *Guide pour la description des sols*. INRA Editions.

[Bharathi Devi et Sarma, 1985] Bharathi Devi B. et Sarma V.V.S., 1985. *Estimation of fuzzy membership histograms*. Information sciences, Vol. 35, 1985, pp. 43-59.

[Bohanec et al., 2013] Bohanec M., Rajkovic V., Bratko I., Zupan B., Znidarsic M., 2013. *DEX methodology: Three decades of qualitative multi-attribute modelling*. Informatica 37, 49-54, 2013.

[Bouchon-Meunier, 1993] Bouchon-Meunier B., 1993. *La logique floue*. PUF « Que Sais-Je ? ». 4ème édition (mise à jour 2007).

[Brans et Mareschal, 2002] Brans J.P. et Mareschal B., 2002. *PROMETHEE-GAIA. Une Méthodologie d'Aide à la Décision en Présence de Critères Multiples*. Ellipses, Paris, France, 2002.

[Brisson et al., 1998] Brisson N., Mary B., Ripoche D., Jeuffroy M.H., Ruget F., Gate P., Devienne-Barret F., Antonioletti R., Durr C., Nicoullaud B., Richard G., Beaudoin N., Recous S., Tayot X., Plenet D., Cellier P., Machet J.M., Meynard J.M., Delecolle R., 1998. *STICS: a generic model for the simulation of crops and their water and nitrogen balance. I. Theory and parametrization applied to wheat and corn*. *Agronomie* 18, 311-346.

[Brisson et al., 2002] Brisson N., Ruget F., Gate P., Lorgeou J., Nicoullaud B., Tayot X., Plenet D., Jeuffroy M.H., Bouthier A., Ripoche D., Mary B., Justes E., 2002. *STICS: a generic model for the simulation of crops and their water and nitrogen balances. II. Model validation for wheat and corn*. *Agronomie*, 22, 69-93.

[Brisson et al., 2003] Brisson N., Gary C., Justes E., Roche R., Mary B., Ripoche D., Zimmer D., Sierra J., Bertuzzi P., Burger P., Bussiere F., Cabidoche Y.M., Cellier P., Debaeke P., Gaudiellere J.P., Maraux F., Seguin B., Sinoquet H., 2003. *An overview of the crop model STICS*. *Eur. J. Agron.* 18, 309-332.

[Choquet, 1953] Choquet, G., 1953. *Theory of capacities*. *Annales de l'institut Fourier* 5 (1954): 131-295.

[Coupland et John, 2008] Coupland S., John R., 2008. *New geometric inference techniques for type-2 fuzzy sets*. *International Journal of Approximate Reasoning*, Volume 49, Issue 1, September 2008, Pages 198-211, ISSN 0888-613X, <http://dx.doi.org/10.1016/j.ijar.2008.03.001>

[Craheix et al., 2012] Craheix D., Angevin F., Bergez J.E., Bockstaller C., Colomb B., Guichard L., Reau R., Doré T., 2012. *MASC 2.0, un outil d'évaluation multicritère pour estimer la contribution des systèmes de culture au développement durable*. *Innovations Agronomiques* 20 (2012), 35-48.

[Dantan et al., 2015 (1)] Dantan, J., Pollet, Y., Taïbi, S., 2015. A systemic meta-model for socio-environmental systems. In proceedings of the Sixth International Conference on Complex Systems Design & Management, CSD&M 2015. Editors: Auvray, G., Bocquet, J.-C., Bonjour, E., Krob, D. (Eds.). November 23-25, 2015. P. 307. Paris, France.

[Dantan et al., 2015 (2)] Dantan, J., Pollet, Y., Taïbi, S. 2015. *Combination of imperfect data in fuzzy and probabilistic extension classes*. Journal of Environmental Accounting and Management. 3 (2), 123-150. DOI: 10.5890/JEAM.2015.06.004.

[Dantan et al., 2014 (1)] Dantan, J., Pollet, Y., Taïbi, S. 2014 (1). *A goal-oriented meta-model for scientific research*. In proceedings of Computational Science and Its Applications - ICCSA 2014 - 14th International Conference, Part V, Lecture Notes in Computer Science LNCS 8583, ISBN 978-3-319-09155-6, pp. 762–774. Springer International Publishing Switzerland. Guimarães, Portugal, June 30 - July 3, 2014.

[Dantan et al., 2014 (2)] Dantan, J., Pollet, Y., Taïbi, S. 2014 (2). *Taking account of uncertain, imprecise and incomplete data in sustainability assessments in agriculture*. In proceedings of Computational Science and Its Applications - ICCSA 2014 - 14th International Conference, Part III, Lecture Notes in Computer Science LNCS 8581, ISBN 978-3-319-09149-5, pp. 625–639. Springer International Publishing Switzerland. Guimarães, Portugal, June 30 - July 3, 2014.

[Dantan et al., 2012 (1)] Dantan J., Pollet Y., Taïbi S., 2012. *Semantic indexation of Web services for collaborative expert activities*. In proceedings of IADIS International Conference Information Systems. Berlin, Germany. March 10-12, 2012. pp. 57-64.

[Dantan et al., 2012 (2)] Dantan, J., Pollet, Y., Taïbi, S., 2012. *A KDD Process to retrieve and aggregate data from relational databases*. In proceedings of IADIS International Conference Information Systems. March 10-12, 2012, pp. 443-445, Berlin, Germany.

[Dempster, 1967] Dempster A., 1967. *Upper and lower probabilities induced by multivalued mapping*. Annals of Mathematical Statistics, AMS-38:325–339.

[Denœux, 2010]. Denœux T., 2010. *Théorie des fonctions de croyance : application en reconnaissance de formes et en fusion d'informations*. Journée « Théorie et applications des fonctions de croyance pour les systèmes de vision », Groupe de travail SCATI, GDR ISIS, Paris, 9 décembre 2010.

[De Rosnay, 1975] De Rosnay J., 1975. *Le macroscope*. Éditions du seuil. ISBN : 978-2-02-004567-4.

[Destercke et al., 2007] Destercke S., Dubois D., Chojnacki E., 2007. *On the relationships between random sets, possibility distributions, p-boxes and clouds*. 28th Linz Seminar on fuzzy set theory, 2007, Linz, Austria.

[Destercke et al., 2008] Destercke S., Dubois D., Chojnacki E., 2008. *Unifying practical uncertainty representations – I: Generalized p-boxes*. International Journal of Approximate Reasoning, Volume 49, Issue 3, November 2008, Pages 649-663, ISSN 0888-613X, <http://dx.doi.org/10.1016/j.ijar.2008.07.003>.

[Destercke et Dubois, 2009] Destercke S., Dubois D., 2009. *The role of generalised p-boxes in imprecise probability models*. In proceedings of 6. International Symposium on Imprecise Probability (p. 179-188). Presented at ISIPTA '09, Durham, GBR (2009-07-14 - 2009-07-18).

[Doré et al., 2006] Doré T., Le Bail M., Martin P., Ney B., Estrade J.R., 2006. *L'agronomie aujourd'hui*. Quae éditions, 2006. ISBN : 978-2759200009.

[Dubois et Prade, 1980] Dubois D. et Prade H., 1980. *Fuzzy Sets and Systems :Theory and Application*. Mathematics in Science and Engineering, Vol.144. Academic Press, New York-London, 1980.

[Dubois et Prade, 1983] Dubois D. et Prade H., 1983. *Unfair coins and necessity measures : toward a possibilistic interpretation of histograms*. Fuzzy Sets and Systems, 10, 1983, pp. 15-20.

[Dubois et Prade, 1988] Dubois D., Prade H., 1988. *Théorie des possibilités, Application à la représentation des connaissances en informatique*. Masson 1988.

[Dury, 2011] Dury J., 2011. *The cropping-plan decision-making : a farm level modelling and simulation approach*. Page 111. Thèse de doctorat. INP (Institut National Polytechnique) de Toulouse.

[ECMA, 2013] ECMA (European association for standardizing information and communication systems), 2013. *The JSON Data Interchange Format. ECMA Standard 404*.

[En ligne]. 1<sup>st</sup> edition / October 2013, consulté le 01/03/2016, <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>

[Edwards, 1971] Edwards W., 1971. *Social utilities*. Engineering Economist, Summer Symposium Series, 6 : 119–129, 1971.

- [Espinasse, 2010] Espinasse B., 2010. *Méthodes et outils pour l'aide à la décision*. [En ligne]. Cours de l'École Polytechnique Universitaire de Marseille, consulté le 01/03/2016, [http://www.lsis.org/espinasseb/Syllabus/syllabus\\_moad.html](http://www.lsis.org/espinasseb/Syllabus/syllabus_moad.html)
- [FAO, 2015] Food and Agriculture Organization of the United Nations (FAO), 2015. *AGROVOC*. [En ligne], consulté le 01/03/2016, <http://aims.fao.org/fr/agrovoc>
- [Fayyad et al., 1996 (1)] Fayyad U. et al., 1996. *From Data Mining to Knowledge Discovery in Databases*. [en ligne], consulté le 01/03/2016, <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>
- [Fayyad et al., 1996 (2)] Fayyad et al., 1996. *KDD for science Data Analysis: issues and examples*. [En ligne], consulté le 01/03/2016, <http://cbse.soe.ucsc.edu/sites/default/files/fayyad96kdd.pdf>
- [Ferson et Ginzburg, 1996] Ferson, S. and Ginzburg, L. R. , 1996. *Different methods are needed to propagate ignorance and variability*. Reliability Engineering and System Safety 54, 133–144.
- [Ferson et al., 2003] Ferson S., Kreinovich V., Ginzburg L., Myers D.S., Sentz K., 2003. *Constructing probability boxes and Dempster–Shafer structures*. [En ligne], Sandia National Laboratories, Albuquerque, NM, SAND2002-4015, 2003, consulté le 01/03/2016, <http://www.sandia.gov/epistemic/Reports/SAND2002-4015.pdf>
- [Fielding, 2000] Fielding R. T., 2000. *Architectural Styles and the Design of Network-based Software Architectures*. [En ligne]. Thèse de doctorat de l'université de Californie, consulté le 01/03/2016, [https://www.ics.uci.edu/~fielding/pubs/dissertation/rest\\_arch\\_style.htm](https://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm)
- [FRCA Centre, 2004] FRCA Centre, 2004. *Diage : Diagnostic Global d'exploitation*. (Fédération Régionale des Coopératives Agricoles du Centre).
- [Gacogne, 1997] Gacogne L., 1997. *Éléments de logique floue*. CNAM, Institut d'informatique d'Entreprise, p. 47, may 1997.
- [Goguen, 1967] Goguen J.A., 1967. *L-fuzzy sets*, *Journal of Mathematical Analysis and Applications*. Volume 18, Issue 1, April 1967, Pages 145-174, ISSN 0022-247X, [http://dx.doi.org/10.1016/0022-247X\(67\)90189-8](http://dx.doi.org/10.1016/0022-247X(67)90189-8).
- [Gong et al., 2015] Gong Y., Hu N., Zhang J., Liu G., Deng J., 2015. *Multi-attribute group decision making method based on geometric Bonferroni mean operator of trapezoidal*

*interval type-2 fuzzy numbers*. Computers & Industrial Engineering, Volume 81, March 2015, Pages 167-176, ISSN 0360-8352, <http://dx.doi.org/10.1016/j.cie.2014.12.030>

[Guillard et al., 2011] Guillard V., Guillaume C., Destercke S., 2012. *Parameters uncertainties and error propagation in modified atmosphere packaging modelling*. Postharvest Biology and Technology, Elsevier, 2012, 67, pp.154-166.

[Hénin et Dupuis, 1945] Hénin S., Dupuis M., 1945. *Essai de bilan de la matière organique du sol*. Ann. Agro 1, 6-26.

[INRA, 2002] INRA (Institut National de la Recherche Agronomique), 2002. *La méthode Indigo*. [En ligne]. UMR Nancy-Colmar Equipe Agriculture Durable, consulté le 01/03/2016, <http://www.inra.fr/indigo/fra/methode.html>

[Klement et al., 1986] Klement, E.P., Puri, M.L., Ralescu, D.A., 1986. *Limit theorems for fuzzy random variables*. Proceedings of The Royal Society of London A 407, 171-182.

[Klir et Parviz, 1992] Klir G.J. et Parviz B., 1992. *Probability-possibility transformations: a comparison*. Intern. J. of General Systems, Vol 21, 1992, pp. 291-310.

[Klir et Wierman, 1998] Klir G.J. et Wierman M.J., 1998. *Uncertainty-based information*. Elements of generalized information theory, Studies in Fuzziness and Soft-Computing. Physica-Verlag, 1998, 168 pages.

[Kwakernaak, 1978] Kwakernaak, H. 1978. *Fuzzy Random Variables—I. Definitions and Theorems*. Information Sciences 15(1), 1-29.

[Kwakernaak, 1979] Kwakernaak, H. 1979. *Fuzzy Random Variables—II. Algorithms and Examples for the Discrete Case*. Information Sciences 17(3), 253-278.

[Lasserre, 1999] Lasserre V., 1999. *Modélisation floue des incertitudes de mesures de capteurs*. Thèse de doctorat dirigée par Foulloy, Laurent et Mauris, Gilles Électronique, électrotechnique, automatique Chambéry 1999, 1 vol. (180 p.).

[Le Moigne, 1977] Le Moigne J-L., 1977. *La théorie du système général*. PUF.

[Liang et Mendel, 2000] Liang Q., Mendel J. M., 2000. *Interval type-2 fuzzy logic systems: Theory and design*. IEEE Trans. Fuzzy Syst., vol. 8, no. 5, pp.535 -550 2000.

- [Liao et al., 2012] Liao S-H, Chu P-H, Hsiao P-Y, 2012. *Data mining techniques and applications – A decade review from 2000 to 2011*. Expert Systems with Applications. Volume 39, Issue 12, 15 September 2012, Pages 11303–11311.
- [Liu et Liu, 2003] Liu, Y.-K., Liu, B., 2003. *Fuzzy Random Variables: A Scalar Expected Value Operator*. Fuzzy Optimization and Decision Making 2(2), 143-160.
- [Magnan de Bornier, 2015] Magnan de Bornier, 2015. *Chapitre 3 - les décisions de production*. [En ligne]. Éléments d'introduction à l'analyse économique, consulté le 01/03/2016, <http://junon.univ-cezanne.fr/bornier/ecoge.html>
- [Makowski, 2005] Makowski, D. 2005. *Estimation de paramètres - principes généraux*. [En ligne], Formation INRA ACTA ICTA, 28 nov. – 1 déc. 2005, consulté le 01/03/2016, [http://www.modelia.org/html/050929\\_formationIntroModDec2005/supportsDesInterventions/MakowskiDcomment\\_Intervention1\\_InraActa.pdf](http://www.modelia.org/html/050929_formationIntroModDec2005/supportsDesInterventions/MakowskiDcomment_Intervention1_InraActa.pdf)
- [Mamdani, 1974] Mamdani, E. H., 1974. *Application of fuzzy algorithms for the control of a simple dynamic plant*. In Proc IEEE (1974), 121-159.
- [Marichal, 2009] Marichal J-L., 2009. *Aggregation functions for decision making*. Decision-Making Process - Concepts and Methods, Wiley 2009.
- [Marshall, 1890] Marshall A., 1890. *Principles of Economics*. 8th edition 1920, MacMillan and co.
- [Masson, 2005] Masson M.-H., 2005. Apports de la théorie des possibilités et des fonctions de croyance à l'analyse de données imprécises. Habilitation à diriger des Recherches soutenue le 2 décembre 2005, Université de technologie de Compiègne.
- [Mendel et John, 2002] Mendel J. M., John R. I., 2002. *Type-2 Fuzzy Sets Made Simple*. IEEE Trans. on Fuzzy Systems, vol. 10, pp. 117–127, April 2002.
- [Meynard et al., 2001] Meynard J.-M., Doré T., Lucas P., 2003. *L'évaluation et la conception de systèmes de culture pour une agriculture durable*. C.R. Acad. Agric. Fr., 87 (4) : 223-236.
- [Moore, 1966] Moore R., 1966. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, N.J., 1966.
- [Moore, 1979] Moore, R., 1979. *Methods and Applications of Interval Analysis*. SIAM Studies in Applied Mathematics. SIAM, Philadelphia.

[OMG, 2015] Object Management Group, 2015. *OMG Systems Modeling Language*. [En ligne], consulté le 01/03/2016, <http://www.omgsysml.org/>

[Ossard et al., 2009] Ossard A., Galan M.B., Boizard H., Leclercq C., Lemoine C., 2009. *Évaluation des impacts environnementaux des pratiques agricoles à l'échelle de la parcelle et de l'exploitation en vue de l'élaboration d'un plan d'actions : une méthode de diagnostic spatialisé fondée sur des indicateurs, le DAEG*. Ingénieries N°59-60, p71-87.

[Palm, 1994] Palm R., 1994. *Processing of signal distributions in control loops*. EUFIT'94, Aachen, Germany, 1994, pp. 310-321.

[Penot et Deheuvels, 2007] Penot E., Deheuvels O., 2007. *Modélisation économique des exploitations agricoles : modélisation, simulation et aide à la décision avec le logiciel Olympe*. Paris : L'Harmattan, p. 9-21.

[Pivert et Prade, 2012] Pivert O., Prade H., 2012. *Détecter les réponses suspectes en présence de données inconsistantes*. Journées Bases de Données Avancées (BDA'12), Oct 2012, Clermont-Ferrand, France. 2012. <hal-00736373>

[Pollet et Robidou, 1998] Pollet Y. et Robidou S., 1998. *SAGESSE : une approche formelle pour la représentation de données dans la Fusion de Données Symboliques*. 7ème Conférence Internationale sur le traitement d'informations et la gestion d'incertitudes dans les systèmes à base de connaissance (IPMU). Paris, juillet 1998. (In French).

[Pollet, 2015] Pollet Y., 2015. *Le langage SysML*. Cours de la chaire d'intégration des systèmes. CNAM Paris.

[Prade et Testemale, 1986] Prade H. et Testemale C., 1986. *Representation of Soft Constraints and Fuzzy Attribute values by means of possibility distributions in databases*. Analysis of Fuzzy Information Vol II, CRC Press 1986.

[Puri et Ralescu, 1986] Puri M.L., Ralescu D.A., 1986. *Fuzzy random variables*. Journal of Mathematical Analysis and Applications 114, 409-422.

[RAD, 2009] RAD (Réseau Agriculture Durable), 2009. *Diagnostic PRAIRIE, outil de diagnostic et de simulation énergétique et économique*. [En ligne], consulté le 01/03/2016, <http://www.adage35.org/diagnostic-changement-de-systeme-prairie/>



- [RAD, 2010] RAD (Réseau Agriculture Durable), 2010. *Diagnostic de durabilité du RAD*. [En ligne], consulté le 01/03/2016, <http://www.agriculture-durable.org/wp-content/uploads/2010/10/Guide-utilisateur-20101.pdf>
- [Raiffa, 1969] Raiffa H., 1969. *Preferences for multi-attributed alternatives*. Technical Report RM-58-68-DOT/RC, The Rand Corporation, Santa Monica, Californie, 1969.
- [Ree et Jager, 1993] Ree R.V. et Jager R., 1993. *Control-data representation at knowledge level*. IEEE Intern. Conf. on Syst., Man and Cybernetics, Le Touquet, France, Vol. 4, 1993, pp. 702-706.
- [Rolland, 2013] Rolland A., 2013. *Méthode en agrégation de préférence*. [En ligne]. Séminaire « agrégation de préférence », laboratoire ERIC, 21 janvier 2013, consulté le 01/03/2016, [http://eric.univ-lyon2.fr/~arolland/supports/Agregation\\_pref\\_rolland.pdf](http://eric.univ-lyon2.fr/~arolland/supports/Agregation_pref_rolland.pdf)
- [Roy, 1968] Roy B., 1968. *Classement et choix en présence de points de vue multiples (la méthode ELECTRE)*. RIRO, 2<sup>e</sup> année, n° 8, 1968, 57-75.
- [Roy, 1996] Roy B., 1996. *Multicriteria Methodology for Decision Analysis*. Kluwer Academic Publishers, 1996.
- [Sadok et al., 2009] Sadok W., Angevin F., Bergez J.E., Bockstaller C., Colomb B., Guichard L., Reau R., Messéan A., Doré T., 2009. *MASC: a qualitative multi attribute decision model for ex ante assessment of the sustainability of cropping systems*. Agron. Sustain. Dev. 29, 447-461.
- [Schvartz et al., 2005] Schvartz C., Muller J.-C., Decroux J., 2005. *Guide de la fertilisation raisonnée*. ISBN : 978-2855571201. COMIFER. Editions France Agricole.
- [Shafer, 1976] Shafer G., 1976. *A mathematical theory of evidence*. Princeton University Press.
- [Shapiro, 2012] Shapiro A.F., 2012. *Implementing Fuzzy Random Variables*. In proceedings of ARCH 2013.1. Society of Actuaries. University of Manitoba, Winnipeg, MB, Canada, August 1-4, 2012.
- [Smets et Kennes, 1994] Smets Ph. and Kennes R., 1994. *The transferable belief model*. Artificial Intelligence 66 (1994) 191-234.
- [Solagro, 1999] Solagro, 1999. *Dialecte : Diagnostic Liant Environnement et CTE*. [En ligne], consulté le 01/03/2016, <http://dialecte.solagro.org/evaluer.php>

[Solagro, 2002] Solagro, 2002. *Dialogue : Diagnostic agri-environnemental de l'exploitation agricole*. [En ligne], consulté le 01/03/2016, [http://www.solagro.org/site/im\\_user/014plaquette\\_dialogue.pdf](http://www.solagro.org/site/im_user/014plaquette_dialogue.pdf)

[Sugeno, 1985] Sugeno M., 1985. *An introductory survey of fuzzy control*. Information and science n°36 p.59, 1985.

[Troffaes et Destercke, 2011] Troffaes M., Destercke S., 2011. *Probability boxes on totally preordered spaces for multivariate modelling*. International Journal of Approximate Reasoning, Volume 52, Issue 6, September 2011, Pages 767-791, ISSN 0888-613X, <http://dx.doi.org/10.1016/j.ijar.2011.02.001>.

[Troffaes et al., 2013] Troffaes M., Miranda E., Destercke S., 2013. *On the connection between probability boxes and possibility measures*. Information Sciences, Volume 224, 1 March 2013, Pages 88-108, ISSN 0020-0255, <http://dx.doi.org/10.1016/j.ins.2012.09.033>

[Vilain et al., 2003] Vilain L. (dir) et al., 2003. *La méthode IDEA - Indicateurs de durabilité des exploitations agricoles*. Guide d'utilisation, 3ème édition. Editions Educagri, Dijon, 100 p.

[Von Neumann et Morgenstern, 1944] Von Neumann J. et Morgenstern O., 1944. *Theory of Games and Economic Behaviour*. Princetown University Press, Princetown, New Jersey, 1944.

[Wolff, 1996] Wolff H.P., 1996. *Économie rurale*. [En ligne]. Université de Ouagadougou, FASEG, 3ème édition, 1996, consulté le 01/03/2016, <https://www.uni-hohenheim.de/~hpwolff/pdfs/HPWEcoAgr.pdf>

[W3C, 2008] W3C (World Wide Web Consortium), 2008. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. [En ligne]. W3C Recommendation 26 November 2008, consulté le 01/03/2016, <https://www.w3.org/TR/xml/>

[Yager, 1988] Yager, R. R., 1988. *On ordered weighted averaging aggregation operators in multi-criteria decision making*. IEEE Transactions on Systems, Man and Cybernetics 18, 183–190, 1988.

[Zadeh, 1965] Zadeh L.A., 1965. *Fuzzy Sets*. Information and Control, Vol 8, Academic Press.

[Zadeh, 1975] Zadeh L.A., 1975. *The Concept of a Linguistic Variable and Its Application to Approximate Reasoning–I*. Information Sciences, vol. 8, pp. 199–249.

[Zadeh, 1978] Zadeh L.A., 1978. *Fuzzy Sets as a basis for a Theory of Possibility*. Fuzzy Sets and Systems, 1, 1978.

## Annexes

## Annexe : code source de l'application proposée

Voici le code source de l'application proposée. Remarque : il s'agit ici uniquement d'une API. Les services ne sont donc pas implémentés.

### Fichier `uncertaintyTrace.py`

Ce script définit les fonctions d'introspection permettant de notifier les définitions de classes et de fonctions ainsi que divers indicateurs qui leurs sont relatifs, e.g. les valeurs des paramètres et le nombre d'appels.

```
"""
Author : Jerome Dantan
"""

#####
#                               Introspection decorators                               #
#####

def traceClass(aclass):
    """
    Prints classes activity
    """
    print("\t***INFO*** CLASS {0} has just been defined".format(aclass))
    if(hasattr(aclass, "_necessity")):
        print("NECESSITY ",aclass._necessity)
    #print(aclass._possibility)
    return aclass

def traceFunction(f):
    """
    Prints functions activity
    """
    def wrapper(*args,**kwargs):
        wrapper.count = wrapper.count + 1
        result = f(*args,**kwargs)
        print("\t***INFO*** FUNCTION {0} with parameters {1} {2} has been
used {3}X".format(f.__name__, args, kwargs,wrapper.count))
        return result
    wrapper.count = 0
    return wrapper

#####
#                               Tests                               #
#####

test = True
if test == True:
    @traceClass
    class FuzzyClass():
        @traceFunction
        def __init__(self, p, n) :
            self.possibility = p
            self.necessity = n
```

```
f = FuzzyClass(1,1)
f = FuzzyClass(2,1)
```

### **Fichier uncertaintyClasses.py**

Ce script définit les classes correspondant aux types modélisant l'incertitude suivants : trapézoïde, loi de probabilité symétrique, gaussienne, distribution de possibilités discrètes, distribution de probabilités discrètes, disjonction de gaussiennes, disjonction de trapézoïdes.

Il définit également des fonctions de conversion des différents types de distributions vers d'autres et des fonctions de vérification de type.

```
"""
Author : Jerome Dantan
"""
from uncertaintyTrace import *
from uncertaintyBelief import *

#####
#                                     Tools                                     #
#####

@traceFunction
def isFuzzyDic(d):
    """
    Checks if "dict" values are a normalized possibility distribution
    """
    for k,v in d.items():
        if v>1 or v<0:
            return False
    for k,v in d.items():
        if v==1:
            return True
    return False

@traceFunction
def toFuzzyDic(d):
    """
    Normalizes "dict" values to make them be a possibility distribution
    """
    for k,v in d.items():
        if v>1:
            d[k]=1
        elif v<0:
            d[k]=0
    maxd=0
    for k,v in d.items():
        if v>maxd:
            maxd = v
    for k,v in d.items():
        d[k]=v/maxd
```

```

    return d

@traceFunction
def isProbaDic(d):
    """
    Checks if "dict" values are a probability distribution
    """
    for k,v in d.items():
        if v>1 or v<0:
            return False
    sump=0
    for k,v in d.items():
        sump = sump + v
    if sump != 1:
        return False
    return True

@traceFunction
def toProbaDic(d):
    """
    Transforms "dict" values into a probability distribution
    """
    for k,v in d.items():
        if v>1:
            d[k]=1
        elif v<0:
            d[k]=0
    sump=0
    for k,v in d.items():
        sump = sump + v
    if sump != 1:
        for k,v in d.items():
            d[k]=v/sump
    return d

@traceFunction
def possToDSDic(do):
    """
    Transforms possibility dict values into a Dempster-Shafer distribution
    """
    d = dict(do)
    n=len(d)
    i=0
    f = FocalElementList()
    #oldfe = ""
    if n>1:
        n=len(d)
        while i<=n-1:
            maxposs = max(d.values())
            if i==0:
                maxposs_old = maxposs
                print("maxposs",maxposs)
                item = ""
                for k,v in d.items():
                    if v==maxposs:
                        item=k
            if(i==0):
                item_old = item
                old_set = set([item_old])
                print("old_set",old_set,maxposs_old,maxposs)

```

```

else:
    print("old_set",old_set,maxposs_old,maxposs)
    fe = FocalElement(old_set, maxposs_old-maxposs)
    old_set = old_set.union(set([item]))
    fe.printFocalElement()
    f.append(fe)
if i==n-1 and len(d)==1:
    for k,v in d.items():
        old_set = old_set.union(set([k]))
        fe = FocalElement(old_set, v)
        fe.printFocalElement()
        f.append(fe)
    i=i+1
    maxposs_old = maxposs
    item_old = item
    del d[item_old]
return f

@traceFunction
def probaToDSDic(do):
    """
    Transforms proba dict values into a Dempster-Shafer distribution
    """
    d = dict(do)
    f = FocalElementList()
    #oldfe = ""
    for k,v in d.items():
        fe = FocalElement(set([k]), v)
        fe.printFocalElement()
        f.append(fe)
    return f

#####
#                               Uncertainty classes                               #
#####

@traceClass
class Trapezoid():
    """
    Trapezoid centered on c, with kernel (c-k,c+k) and support (c-s,c+s)
    """
    def __init__(self, c, k, s):
        self.central = c
        self.kernel = k
        self.support = s
    def printU(self):
        print(self, self.central, self.kernel, self.support)

@traceClass
class VA():
    """
    Random variable of any law with mode "a" and variance "v"
    """
    def __init__(self, a, v):
        self.central = a
        self.variance = v
    def printU(self):
        print(self.central, self.variance)

@traceClass

```



```

class Normal():
    """
    Gaussian random variable with mode "a" and variance "v"
    """
    def __init__(self, a, v) :
        self.central = a
        self.variance = v
    def printU(self):
        print(self.central, self.variance)

@traceClass
class FuzzyDiscrete():
    """
    Discrete possibility distribution
    """
    def __init__(self, d) :
        if not(isFuzzyDic(d)):
            self.dic = toFuzzyDic(d)
        else:
            self.dic = d
    def printU(self):
        print(self.dic)

    def getDic(self):
        return self.dic

@traceClass
class ProbaDiscrete():
    """
    Discrete probability distribution
    """
    def __init__(self, d) :
        if not(isProbaDic(d)):
            self.dic = toProbaDic(d)
        else:
            self.dic = d
            self.dic = d
    def printU(self):
        print(self.dic)

def fromPossibilityDicToBBA(fd):
    d = fd.getDic()
    for k,v in d.items():
        print(k)
        print(v)

@traceClass
class ProbaNormalCollection():
    """
    Set of independant Gaussian random variables
    """
    def __init__(self, d):
        if not(isProbaDic(d)):
            self.dic = toProbaDic(d)
        else:
            self.dic = d
            self.dic = d
    def printU(self):
        print(self.dic)

```

```

@traceClass
class FuzzyTrapezoidCollection():
    """
    Discrete set of trapezoid distributions
    """
    def __init__(self, d) :
        if not(isFuzzyDic(d)):
            self.dic = toFuzzyDic(d)
        else:
            self.dic = d
    def printU(self):
        print(self.dic)

#####
#                                     Tests                                     #
#####
fdis=FuzzyDiscrete({3:1,8:0.7,2:0.2,4:0.1})
pdis=ProbaDiscrete({3:0.2,8:0.5,2:0.1,4:0.2})

f = possToDSDic(fdis.dic)
p = probaToDSDic(pdis.dic)
f.printFocalElementList()
p.printFocalElementList()
print(f.isPossibilityDistribution())
print(f.isProbabilityDistribution())
print(f.isPossibilityDistribution())
print(f.isProbabilityDistribution())

print("***** ORTHOGONAL * *****")
result = f * p
result.printFocalElementList()
conjonctivefEltList.printFocalElementList()

```

## **Fichier uncertaintyTools.py**

Ce script définit les outils de calcul de l'imperfection des données. Il définit fonctions de calculs algébriques sur les types modélisant l'incertitude définis dans le fichier uncertaintyClasses.py. Il définit également les métaclasses permettant de définir des classes dynamiquement ainsi que l'annotation « uncertainty » qui permet d'annoter les attributs de classes avec les différents types d'incertitude.

```

"""
Author : Jerome Dantan
"""
from uncertaintyClasses import *
import operator

# Root operator list
ops = { "+": operator.add, "-": operator.sub, "*": operator.mul, "/":
operator.truediv }

```

```
#####
#                                     General sum                                     #
#####

@traceFunction
def hybridOperation(u1, u2, op):
    if isinstance(u2, Trapezoid): # Trapezoid
        if u2.central==0 and op=="/":
            print("WARNING, division per 0. Intersection applied")
            return intersection(u1, u2)
        else:
            return Trapezoid(ops[op](u1, u2.central), u2.kernel, u2.support)
    elif isinstance(u2, Normal): # Normal
        return Normal(ops[op](u1, u2.central), u2.variance)
    elif isinstance(u2, VA): # any VA
        return VA(ops[op](u1, u2.central), u2.variance)
    elif isinstance(u2, FuzzyDiscrete): # FuzzyDiscrete
        u3={}
        for k,v in u2.dic.items():
            u3[operation(k,u1,op)]=float(v)
        return FuzzyDiscrete(u3)
    elif isinstance(u2, ProbaDiscrete): # ProbaDiscrete
        u3={}
        for k,v in u2.dic.items():
            u3[operation(k,u1,op)]=float(v)
        return ProbaDiscrete(u3)
    elif isinstance(u2, ProbaNormalCollection): # ProbaNormalCollection
        u3={}
        for k,v in u2.dic.items():
            u3[operation(k,u1,op)]=float(v)
        return ProbaNormalCollection(u3)
    elif isinstance(u2, FuzzyTrapezoidCollection): #
FuzzyTrapezoidCollection
        u3={}
        for k,v in u2.dic.items():
            u3[operation(k,u1,op)]=float(v)
        return FuzzyTrapezoidCollection(u3)
    else:
        print(type(u1), u1)
        print(type(u2), u2)
        print("WARNING, function not implemented. Intersection applied")
        return intersection(u1, u2)

@traceFunction
def operation(u1, u2, op):
    # Primitive types
    if (isinstance(u1,int) or isinstance(u1,float)) and (isinstance(u2,int)
or isinstance(u2,float)):
        return ops[op](u1, u2)
    # Primitive types + others
    elif isinstance(u1,int) or isinstance(u1,float):
        return hybridOperation(u1,u2,op)
    elif isinstance(u2,int) or isinstance(u2,float):
        return hybridOperation(u2,u1,op)
    # Others
    elif (isinstance(u1,Trapezoid) and isinstance(u2,Trapezoid)): #
Trapezoid
        if u2.central==0 and op=="/":
            print("WARNING, division per 0. Intersection applied")
            return intersection(u1, u2)
```

```

        else:
            central = ops[op](u1.central, u2.central)
            if op=="+" or op=="-":
                return
Trapezoid(central,u1.kernel+u2.kernel,u1.support+u2.support)
            elif op=="*":
                return
Trapezoid(central,u2.central*u1.kernel+u1.central*u2.kernel,u2.central*u1.s
upport+u1.central*u2.support)
            elif op=="/":
                return
Trapezoid(central, (u2.central*u1.kernel+u1.central*u2.kernel)/u2.central**2
, (u2.central*u1.support+u1.central*u2.support)**2)
            elif(isinstance(u1,Normal) and isinstance(u2,Normal)): # Normal
                central = ops[op](u1.central, u2.central)
                if op=="+" or op=="-":
                    return Normal(central,u1.variance+u2.variance)
                elif op=="*":
                    return
VA(central, (u1.variance**2+u1.central**2)*(u2.variance**2+u2.central**2)**2
-u1.central**2*u2.central**2)
                elif op=="/":
                    print("WARNING, function not implemented. Intersection
applied")
                    return intersection(u1, u2)
            elif(isinstance(u1,VA) and isinstance(u2,VA)): # any VA
                central = ops[op](u1.central, u2.central)
                if op=="+" or op=="-":
                    return VA(central,u1.variance+u2.variance)
                elif op=="*":
                    return
VA(central, (u1.variance**2+u1.central**2)*(u2.variance**2+u2.central**2)**2
-u1.central**2*u2.central**2)
                elif op=="/":
                    print("WARNING, function not implemented. Intersection
applied")
                    return intersection(u1, u2)
            elif(isinstance(u1,FuzzyDiscrete) and isinstance(u2,FuzzyDiscrete)): #
FuzzyDiscrete
                u3={}
                for key1,value1 in u1.dic.items():
                    for key2,value2 in u2.dic.items():

u3[operation(key1,key2,op)]=min(float(value1),float(value2))
                return FuzzyDiscrete(u3)
            elif(isinstance(u1,ProbaDiscrete) and isinstance(u2,ProbaDiscrete)): #
ProbaDiscrete
                u3={}
                for key1,value1 in u1.dic.items():
                    for key2,value2 in u2.dic.items():
                        u3[operation(key1,key2,op)]=float(value1)*float(value2)
                return ProbaDiscrete(u3)
            elif(isinstance(u1,ProbaNormalCollection) and
isinstance(u2,ProbaNormalCollection)): # ProbaNormalCollection
                u3={}
                for key1,value1 in u1.dic.items():
                    for key2,value2 in u2.dic.items():
                        u3[operation(key1,key2,op)]=float(value1)*float(value2)
                return ProbaNormalCollection(u3)

```

```

        elif(isinstance(u1,FuzzyTrapezoidCollection) and
isinstance(u2,FuzzyTrapezoidCollection)): # FuzzyTrapezoidCollection
            u3={}
            for key1,value1 in u1.dic.items():
                for key2,value2 in u2.dic.items():

u3[operation(key1,key2,op)]=min(float(value1),float(value2))
            return FuzzyTrapezoidCollection(u3)
        else:
            print(type(u1), u1)
            print(type(u2), u2)
            print("WARNING, function not implemented. Intersection applied")
            return intersection(u1, u2)

@traceFunction
def addition(u1,u2):
    return operation(u1,u2,"+")

@traceFunction
def multiplication(u1,u2):
    return operation(u1,u2,"*")

@traceFunction
def subtraction(u1,u2):
    return operation(u1,u2,"-")

@traceFunction
def division(u1,u2):
    return operation(u1,u2,"/")

@traceFunction
def intersection(u1, u2):
    if(isinstance(u1,FuzzyDiscrete) and isinstance(u2,FuzzyDiscrete)):
        u3={}
        for key1,value1 in u1.dic.items():
            for key2,value2 in u2.dic.items():
                u3[key1, key2]=min(float(value1),float(value2))
        return FuzzyDiscrete(u3)
    else:
        return [u1, u2]

#####
#                                     Meta classes                                     #
#####

@traceClass
class metaclass(type):
    def __new__(metaccls, name, bases, dict):
        dict["_possibility"]=1
        dict["_necessity"]=1
        return type.__new__(metaccls, name, bases, dict)
    def __init__(metaccls, name, bases, dict):
        """meta class constructor (called when creating a new class)"""
        return type.__init__(metaccls, name, bases, dict)

#####
#                                     Decorators                                     #
#####

@traceFunction

```

```

def uncertainty(attributes):
    def decorator(f):
        print("Decorator with following parameters : ", str(attributes))
        def wrapped(param1,param2, param3):
            print("param1 : ",param1)
            print("param2 : ",param2)
            print("param3 : ",param3)
            for att in attributes:
                if att["attribute"]==param2:
                    if att["type"]=="trapezoid":
                        param1.__setattr__(param2,Trapezoid(param3,
att["kernel"], att["support"]))
                    elif att["type"]=="normal":
                        param1.__setattr__(param2,Normal(param3,
att["variance"]))
                    elif att["type"]=="fuzzyDiscrete":
                        param1.__setattr__(param2,FuzzyDiscrete(att["distribution"]))
                    elif att["type"]=="probaDiscrete":
                        param1.__setattr__(param2,ProbaDiscrete(att["distribution"]))
                else:
                    pass
            #param1.__setattr__(param2,param3)
            return wrapped
        return decorator

#####
#                                     Tests                                     #
#####

test = True
if test == True:
    fcn1=FuzzyTrapezoidCollection({Trapezoid(0,1,2):1,
Trapezoid(1,2,3):0.2})
    fcn2=FuzzyTrapezoidCollection({Trapezoid(5,1.5,2):1,
Trapezoid(3,1,3):0.5})
    print("fcn1",fcn1.dic)
    print("fcn2",fcn2.dic)
    fcn3=addition(fcn1,fcn2)
    print("fcn3",fcn3)
    fcn3.printU()

    t1=Trapezoid(0,1,2)
    t2=Trapezoid(1,2,3)
    t=addition(t1,3)
    t.printU()

    n1=Normal(0,1)
    n2=2
    n=addition(n1,n2)
    n.printU()

```

## Fichier Tests.py

Ce script définit une classe de test définissant notamment des accesseurs (set) et mutateurs (get) génériques grâce à des annotations (décorateurs).

```
"""
Author : Jerome Dantan
"""
from uncertaintyTools import *

#####
#                                     Business classes                                     #
#####

defaultKernel = 10
defaultSupport = 20
defaultVariance = 4

@traceFunction
@uncertainty(attributes=[{"attribute": "_crops", "type": "trapezoid",
"kernel":8, "support":15},
                        {"attribute": "_surface", "type": "trapezoid",
"kernel":10, "support":20},
                        {"attribute": "_prod", "type": "normal",
"variance":8},
                        {"attribute": "_fertilization", "type": "normal",
"variance":10},
                        {"attribute": "_size", "type": "fuzzyDiscrete",
"distribution":{2:1, 3:0.2}},
                        {"attribute": "_fert1", "type": "fuzzyDiscrete",
"distribution":{4:1, 6:0.5}}
])
def __set__(self, attribute, value):
    """
    Generic setter
    """
    self.__setattr__(attribute, value)

@traceFunction
def __get__(self, attribute):
    """
    Generic getter
    """
    return self.__getattr__(attribute)

ParcelMethods = {
    "color": ('rouge', 'blanc'),
    "_surface" : 0,
    "_prod" : 0,
    "_size" : 0,
    "_get" : _get,
    "_set" : _set
}

CropProductionMethods = {
    "color": ('rouge', 'blanc'),
```

```

        "_crops" : 0,
        "_fertilization" : 0,
        "_fert1" : 0,
        "_get" : _get,
        "_set" : _set
    }

@traceClass
class Farmer():
    def __init__(self) :
        self._money = 0
        @uncertainty(attributes=[{"attribute": "_money", "type": "trapezoid",
"kernel":1, "support":2}])
    def _set(self, attribute, value):
        """
        Setter
        """
        self.__setattr__(attribute, value)
    @traceFunction
    def _get(self, attribute):
        """
        Generic getter
        """
        return self.__getattr__(attribute)

#####
#                                     Tests                                     #
#####

# Classes
CropProduction = metaclass("CropProduction", (), CropProductionMethods)
Parcel = metaclass("Parcel", (), ParcelMethods)

# Objets
mycp = CropProduction()
myp = Parcel()

f = Farmer()
f._set("_money", 100000)
print(f._get("_money"))
f._get("_money").printU()

T=True
F = False
testT = T
testN = T
testFD = T

if testFD == True:
    print("***** testFD *****")
    myp._set("_size", 10)
    mycp._set("_fert1", 20)
    t = intersection(myp._get("_size"), mycp._get("_fert1"))
    print("FUZZY_DISCRETE ", t)
    t.printU()

if testT == True:
    mycp._set("_crops", 3)
    print("_crops ", mycp._get("_crops"))
    myp._set("_surface", 15)

```



```

print("_surface ", myp._get("_surface"))
t = addition(myp._get("_surface"), mycp._get("_crops"))
print("TRAPEZOID ",t)
t.printU()

if testN == True:
    mycp._set("_fertilization",4)
    print("_fertilization ", mycp._get("_fertilization"))
    myp._set("_prod", 12)
    print("_prod ", myp._get("_prod"))
    s = addition(mycp._get("_fertilization"), myp._get("_prod"))
    print("NORMAL ",s)
    s.printU()

```

# **Une approche systémique unifiée pour l'optimisation durable des systèmes socio-environnementaux**

## **Résumé**

De nos jours, la durabilité des activités humaines devient une préoccupation majeure dans le monde entier. Il s'agit d'évaluer ces activités non seulement en matière d'efficacité et de productivité, mais aussi en ce qui concerne leurs durabilités économique, sociale, environnementale, etc. Pour ce faire, les experts de ces différents domaines doivent travailler en collaboration. Dans ce contexte, les sociétés humaines sont confrontées à plusieurs défis majeurs qui sont les suivants : (1) traiter de grandes quantités d'informations (« big data »), (2) évoluer dans un monde réel dynamique et imparfait, (3) prévoir et évaluer les états futurs de ses activités.

Les recherches que nous avons menées dans cette thèse contribuent plus particulièrement au domaine de l'ingénierie des systèmes de décision en univers incertain. Nous prenons comme objet d'étude général le domaine des systèmes socio-environnementaux, et plus particulièrement le domaine pluridisciplinaire de l'agriculture. Nous proposons une approche systémique pour l'optimisation durable des systèmes socio-environnementaux : (1) la méta-modélisation des systèmes socio-environnementaux, (2) la représentation générique de l'imperfection des informations qui circulent dans ces systèmes, associée à un modèle de décision en contexte incertain et enfin (3) la simulation et l'évaluation de ces systèmes en environnement dynamique en vue de prises de décisions par des experts, que nous avons illustrée par un modèle d'architecture orientée service ainsi que des études de cas appliquées au domaine de l'agriculture.

Mots-clés : décision, incertitude, optimisation durable, possibilité, probabilité, systèmes socio-environnementaux.

## **Résumé en anglais**

Nowadays, the sustainability of human activities is a major worldwide concern. The challenge is to evaluate such activities not only in terms of efficiency and productivity, but also in terms of their economic, social, environmental, etc. durability. For this, the experts of these areas need to work collaboratively. In this context, human societies are facing several major challenges such as: (1) process a large amount of information whose volume increases exponentially ("big data"), (2) live in a both dynamic and imperfect real world, (3) predict and assess future states of its activities.

The researches we have conducted in this thesis contribute in particular to the domain of decision systems engineering under uncertainty. We have chosen the field of general socio-environmental systems as subject of study, particularly the multidisciplinary field of agriculture. We propose a systemic approach for the sustainable optimization of socio-environmental systems: (1) the meta-modeling of socio-environmental systems, (2) the generic representation of data imperfection flowing in such systems, associated to a decision model in uncertain environment and finally (3) the simulation and the assessment of such systems in dynamic environment for the purpose of decision making by experts which we have illustrated by both a service-oriented architecture model and case studies applied to the agriculture domain.

Keywords: decision, uncertainty, sustainable optimization, possibility, probability, socio-environmental system.